STRUCTURAL AND FUNCTIONAL STUDIES OF MYCOTHIOL BIOSYNTHESIS PRECURSOR ENZYME IN *MYCOBACTERIUM TUBERCULOSIS*

A Thesis

by

WAN WEN ZHU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2011

Major Subject: Chemistry

Structural and Functional Studies of Mycothiol Biosynthesis Precursor Enzyme in

Mycobacterium tuberculosis

STRUCTURAL AND FUNCTIONAL STUDIES OF MYCOTHIOL BIOSYNTHESIS PRECURSOR ENZYME IN *MYCOBACTERIUM TUBERCULOSIS*

A Thesis

by

WAN WEN ZHU

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	James C. Sacchettini
Committee Members,	Frank M. Raushel
	Thomas R. Ioerger
Head of Department,	David H. Russell

August 2011

Major Subject: Chemistry

ABSTRACT

Structural and Functional Studies of Mycothiol Biosynthesis Precursor Enzyme in *Mycobacterium tuberculosis*.

(August 2011)

Wan Wen Zhu, B.S., Polytechnic University

Chair of Advisory Committee: Dr. James C. Sacchettini

MshA is a glycosyltransferase that synthesizes the precursor of mycothiol, a low-molecular-weight thiol found exclusively in *Actinomycetes*, including the virulent pathogen *Mycobacterium tuberculosis* (*Mtb*). The structure of MshA from *Mtb* (herein coined as TbMshA) and its complex with uridine diphosphate N-acetyl-glucosamine (UDP-GlcNAc) have been solved to resolutions of 2.32 Å and 2.89 Å respectively. Both structures form two monomers in the asymmetric unit cell and exhibit typical β/α/β Rossmann folds. Upon binding of UDP-GlcNAc, the C-terminal domain of TbMshA undergoes conformational changes in order to interact with UDP-GlcNAc at the binding site. In addition, ligand-bound TbMshA structure enables the identification of critical residues for enzymatic interactions, especially the residue Glu-353 (E353) at the active site that is believed to serve as a nucleophile in the sugar transfer of TbMshA. In order to verify this, a mutant of TbMshA with a single amino acid mutation from glutamate to glutamine at residue 353 is generated. The mutant (E353Q) has shown reduced enzyme activity by more than four-fold compared to the wild-type TbMshA (V_{max} for wild-type

is 0.17 ± 0.02 μM sec^{-1}, whereas V_{max} for E353Q is 0.04 ± 0.01 μM sec^{-1}). The k_{cat}/K_m for wild-type TbMshA (3.5 ± 1.1 * 10^3 M^{-1} sec^{-1}) is an order of magnitude higher than that of the mutant (0.3 ± 0.1 * 10^3 M^{-1} sec^{-1}), indicating the catalytic efficiency is greatly suppressed by the mutation. Mass spectrometry data also reveals that E353Q is unable to form the product of the reaction catalyzed by the wild-type TbMshA. These findings suggest the important role of Glu-353 in the structure and activity of TbMshA.

NOMENCLATURE

CgMshA	MshA from *Corynebacterium glutamicum*
ESI	Electrospray Ionization
E353Q	a mutant of MshA in *Mtb* generated in this thesis
GlcNAc	N-acetyl-glucosamine
GlcNAc-Ins-P	N-acetyl-glucosamine-inositol-phosphate
I-1-P	1L-myo-inositol-1-phosphate
LDH	lactate dehydrogenase
MDR	multiple-drug resistance
MSH	mycothiol
Mtb	*Mycobacterium tuberculosis*
NAD^+	nicotinamide adenine dinucleotide
NADH	nicotinamide adenine dinucleotide reduced
PK	pyruvate kinase
PO_4	phosphate
TB	Tuberculosis
TbMshA	MshA from *Mycobacterium tuberculosis*
UDP	uridine diphosphate
UDP-GlcNAc	uridine diphosphate N-acetyl-glucosamine
WT	wild-type
XDR	extensive-drug resistance

TABLE OF CONTENTS

Page

LIST OF FIGURES

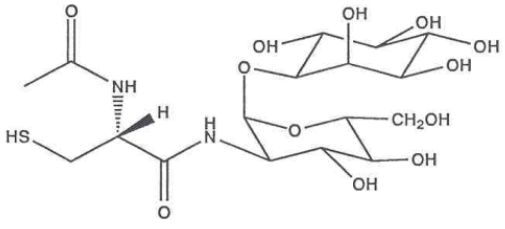
Page

LIST OF TABLES

1. INTRODUCTION

Tuberculosis (TB) is a notorious infectious disease that is affecting nearly one-third of the world's population to date (1). *Mycobacterium tuberculosis* (*Mtb*) is the virulent agent that is responsible for the spread of TB. New cases of TB infection increase daily, especially in poor hygiene and HIV-prevalent countries (1). The occurrence of multiple-drug resistant (MDR) and extensive-drug resistant (XDR) TB also raise the concern of current anti-tubercular regimen (2). In order to improve anti-tubercular treatment at the clinical side, much effort is being paid to understand this deadly pathogen's own defensive mechanisms.

One of the many biosynthetic pathways known for *Mtb* is the mycothiol biosynthesis. Mycothiol (MSH) or 2-(N-acetylcysteinyl) amido-2-deoxy-α-D-glucopyranosyl-(1→1)-myo-inositol (Figure 1) is the major low-molecular weight thiol found in most *Actinomycetes*; it is a class of gram-positive bacteria including the virulent pathogen *Mycobacterium tuberculosis* (*Mtb*) (3).



MYCOTHIOL

Figure 1. The chemical structure of mycothiol (MSH).

This thesis follows the style of *The Journal of Biological Chemistry*.

Bearing the similar antioxidant nature as glutathione (4,5), mycothiol has been found to be conditionally essential for growth in *Mtb* (6), most importantly, mycothiol has not been found in eukaryotes, rendering the enzymes involved in the biosynthetic pathway of mycothiol valuable anti-tuberculosis drug targets.

There are four enzymes that have been identified in mycothiol biosynthesis (Figure 2):

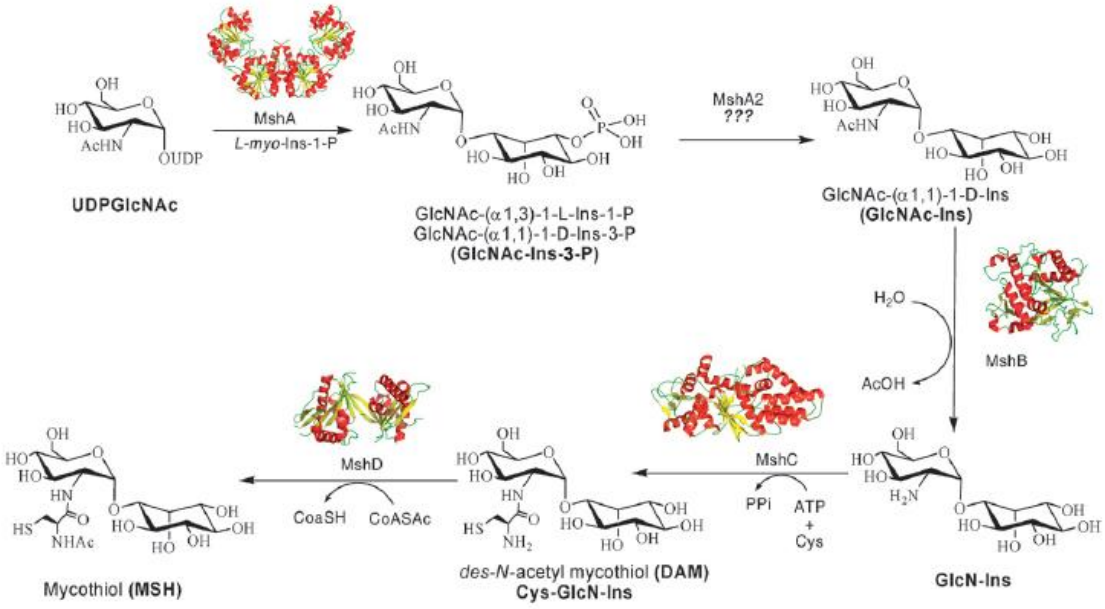


Figure 2. The biosynthetic pathway of mycothiol (7).

MshA is a glycosyltransferase proposed to catalyze the reaction of uridine diphosphate N-acetyl-glucosamine (UDP-GlcNAc) with 1L-myo-inositol-1-phosphate (I-1-P) to generate GlcNac-(α1,3)-1L-Ins-1-P (8) or known as N-acetyl-glucosamine-inositol-phosphate (GlcNAc-Ins-P, in this thesis), which is then deacetylated by MshB to

generate GlcN-Ins (9). MshC helps to ligate the GlcN-Ins with cysteine utilizing the hydrolysis power of ATP to generate Cys-GlcN-Ins (10), and finally MshD catalyzes the acetylation of Cys-GlcN-Ins using acetyl-CoA to produce mycothiol (11).

MshA has been identified as the enzyme that catalyzes the precursor reaction in the mycothiol pathway. It has been considered essential for mycothiol biosynthesis through genetic knockout studies (8). Deletions of MshA alleles cause defective production of mycothiol, and recent report show that mycothiol biosynthesis even has implications in drug resistance TB; an increased susceptibility to ethionamide (second-line TB drug) resistance in *Mtb* with mutations in the MshA gene is observed (12). In order to understand MshA in mycothiol biosynthesis before investigating its implication in *Mtb* drug resistance, the biochemistry of such mechanism should be elucidated. Although this initial step in the mycothiol biosynthetic pathway has been well characterized in the past decade, and the proposed substrates are defined (13), a crystal structure of this enzyme has not yet been determined. Herein, two crystal structures are solved for MshA from *Mtb*, the native enzyme with no ligand bound, and a UDP-GlcNAc bound MshA complex, in which the latter structure locates the substrate binding cavity and provides hints for this precursor reaction in the biosynthesis of mycothiol.

2. MATERIALS AND METHODS

2.1 Cloning, expression and purification of TbMshA

The original MshA gene (rv0486, 480 amino acids in full length) from *Mycobacterium tuberculosis* genome (H37Rv), coined TbMshA, was amplified via Polymerase Chain Reaction (PCR) using Failsafe^{TM} kit (Epicentre) with primers 5'-GGC CCG AAG CTT TCA CGC GCC CAC CCC GC -3' (forward) and 5'- CCG CGG CAT ATG GCA GGT GTG CGG CAC GAT GAC G-3' (reverse). The amplified TbMshA gene was cloned into kanamycin-resistant pET28b(+) vector (Novagen) with restriction sites of Nde1 and HindIII, containing a N-terminal 6X His-tag and a tobacco etch virus (TEV) cleavage sequence. The TbMshA-pET28b(+) plasmid was transformed in BL21DE3 (Invitrogen) expression system. A single colony was picked and inoculated into a 5-ml starter culture (with 50 μg/ml kanamycin), which would be used to inoculate large Luria-Broth (LB) media (2 x 2-liter) and the cell culture (with kanamycin resistant) were grown at 37^{o}C with shaking at 200 revolutions per minute (RPM) until optical density at 600 nm (OD_{600}) reached ~ 0.8-0.9. Cells were induced using isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown at 18^{o}C for 20-24 hours. Cells were then harvested by centrifugation at 4000 RPM for 30 minutes. Cell pellets were resuspended in 2-3 times of pellet volume of lysis buffer (20 mM Tris, pH 8.5, 500 mM sodium chloride (NaCl), 20% glycerol, 5 mM imidazole) at 4^{o}C. Aliquots of DNase (final concentration of 20 μg/ml) and ethylenediaminetetraacetic acid (EDTA)-free protease inhibitors (CalbioChem) solution were added into the cell resuspension just before

lyzing cells. Cell lysis was done using French Press at 1200 pounds per square inch (PSI). Cell lysate was obtained after ultracentrifugation at 15,000 RMP for 50 minutes. Clear lysate (after filtration using 0.22 micron filter from Fischer Scientific) was loaded onto pre-packed Ni-NTA nickel column (GE Healthcare). The column was washed thoroughly with buffer A (20 mM Tris, pH 8.5, 500 mM NaCl, 20% glycerol, 5 mM imidazole, 2 mM β-mercaptoethanol), and eluted with buffer B (buffer A + 500 mM imidazole) in gradient mode (50-500 mM of imidazole) using fast protein liquid chromatography (FPLC from Akta^{TM}). Pure fractions of proteins were eluted at around 150 mM imidazole and pooled for dialysis at 4^{o}C to get rid of excessive imidazole and salt. Dialyzed fractions were concentrated down via ultrafiltration (Amicon) to a concentration of ~ 5-9 mg/ml and store at 4^{o}C (if set up crystal plates immediately, otherwise flash-freeze at -80^{o}C). Removal of N-terminal 6X His-tag was done by incubating the concentrated protein with TEV enzyme (10:1 protein to enzyme in molar ratio) in 20 mM Tris, pH 8.5, 20% glycerol and 100 mM NaCl, and 1 mM dithiothreitol (DTT) at 16^{o}C for overnight (12 hours).

2.2 Crystallization and structure determination of TbMshA

Crystallizations of both apo and ligand-bound TbMshA were set up in a 24-well sitting drop fashion under silicon oil (Hampton Research) at 16^{o}C. Apo TbMshA crystals were obtained overnight with rhombohedral morphology in 100 mM Tris, pH 8.5, and 2.5 M ammonium phosphate. Ligand-bound TbMshA crystals were obtained overnight with diamond-shape (Figure 3) in 100 mM Tris, pH 8.5, 15% polyethylene

glycol (PEG) 4000, and 10 mM uridine diphosphate glucosamine (UDP-GlcNAc) sodium salt (Sigma).

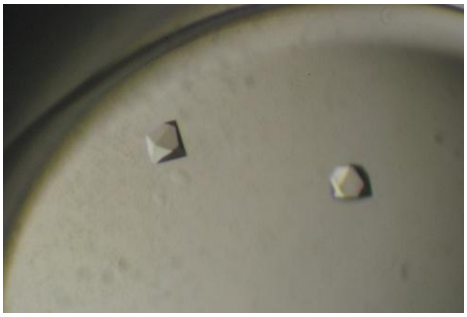


Figure 3. A picture of the UDP-GlcNAc bound TbMshA crystal.

Apo crystal was soaked in crystal condition (mother liquor) with 20% glycerol (final concentration) before mounting in liquid nitrogen at 120 degree Kelvin (K) for data collection. UDP-GlcNAc co-crystallized (ligand-bound) TbMshA crystals were soaked in mother liquor with 25% PEG 400 prior to cryo mounting. Apo TbMshA data (Figure 4) were collected at Advanced Photon Source 23ID-D (Argonne National Laboratory). The initial UDP-GlcNAc bound TbMshA data were collected at 23ID-B and also at home source (soaked with 10 mM myo-inositol and 10 mM ammonium phosphate, final concentration) from Bruker with CCD image detector.

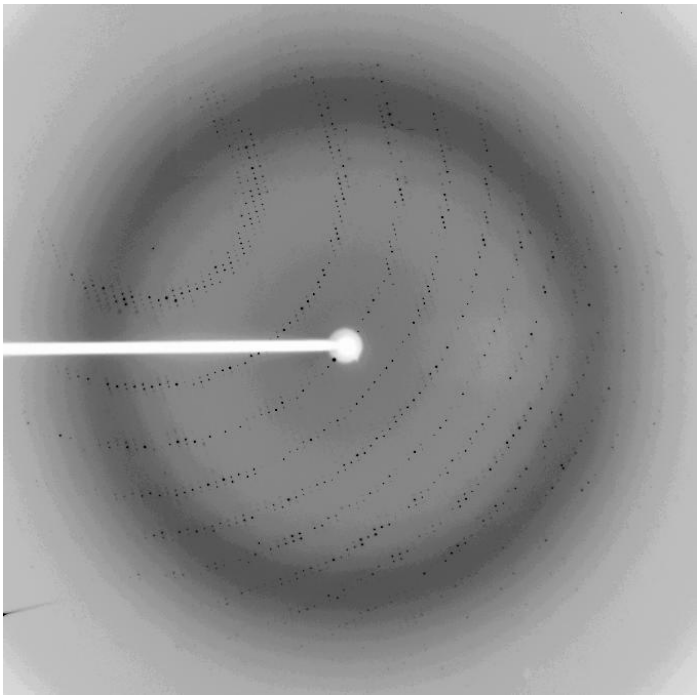


Figure 4. X-ray diffraction of apo TbMshA crystal.

Both apo and ligand-bound data sets were processed using HKL2000. The space group of apo TbMshA protein was determined to be $P2_12_12$ with unit cell dimensions of a = 159.3 Å, b = 56.6 Å, c = 106.6 Å, $\alpha = \beta = \gamma = 90^o$. And the space group of ligand-bound TbMshA was determined to be $P4_12_12$ with unit cell dimensions of a = 123.6 Å, b = 123.6 Å, c = 154.0 Å, $\alpha = \beta = \gamma = 90^o$. Both crystals have two monomers in the asymmetric unit assumed a solvent content of 50%. The structure of UDP-GlcNAc bound TbMshA was determined using molecular replacement (CCP4 molprep) with the solved ligand-bound CgMshA (Protein Data Bank or PDB code: 3c4v) as an input phasing model. The native TbMshA structure was solved by using the C-terminal domain of the ligand-bound TbMshA as the initial searching model, and then used the

searched solution as a fixed searching model while finding another solution using the N-terminal domain as a second search model in CCP4 molrep. Model building was done using COOT for both structures and refined using Phenix. For the UDP-GlcNAc bound TbMshA structure, the ligand PDB was generated using PRODRG server, molecular constraints were calculated using Phenix eLBOW.builder and refined into the structure using Phenix refine. The first 42 residues and the last 29 residues of both TbMshA structures lack electron density after refinement maybe due to high disorderness or flexibilities at the termini.

2.3 Construction of TbMshA mutant (E353Q)

A single amino acid mutation was introduced for TbMshA at residue 353 (a glutamate) via the site-directed mutagenesis kit from Stratagene. In order to construct this mutant, the original TbMshA gene (Rv0486) was PCR amplified using primers 5'-GCG GTG CCG AGC TAC TCC CAG TCG TTC GGC CTG GTT GC -3' (forward) and 5'- GC AAC CAG GCC GAA CGA CTG GGA GTA GCT CGG CAC CGC-3' (reverse), in which the underscored codon was the site where the mutation was introduced. This single amino acid mutation, Q → E, also corresponded to a single nucleotide mutation, G → C (GAG => glutamic acid, CAG => glutamine). The amplified mutant gene (namely E353Q) was cloned into kanamycin-resistant pET28b(+) vector (Novagen) with restriction sites of Nde1 and HindIII, containing a N-terminal 6X His-tag and a TEV cleavage sequence. E353Q-pET28b(+) was transformed into BL21DE3 (Invitrogen) expression system and expressed in the same fashion as that of

original TbMshA gene (or wild-type gene). The mutant TbMshA protein, E353Q, was also purified following the same protocol mentioned previously (Section 2.1) for the wild-type TbMshA.

2.4 Measurement of enzymatic activity for TbMshA

The TbMshA enzyme kinetics assay was carried out in a coupled enzyme assay by monitoring the production of uridine diphosphate (UDP) at 340 nm (14). The standard assay condition contained (final concentration): 200 nM enzyme, 20 mM Tris, pH 8.0, 500 μM NADH, 500 μM phosphoenol pyruvate (PEP), 10 mM magnesium chloride ($MgCl_2$), 8 units of pyruvate kinase (PK in 50% glycerol suspension, from Sigma) and 10 units of lactate dehydrogenase (LDH in 50% glycerol suspension, from Sigma). The final reaction mixture (50 μl) was allowed to equilibrate for 2 minutes before adding the substrates to start the reaction. Reactions were monitored using 96-well flat bottom transparent assay plate (from Greiner) and the results were recorded by 96-well plate reader (Polarstar Omega, BMG labtech).

Donor kinetics was performed for both wild-type and mutant TbMshA, while holding the acceptor substrate (I-1-P) at constant/saturating concentration (2 mM). The concentration of donor substrate (UDP-GlcNAc) varied from 4 mM (final concentration) to 0 mM via serial dilution in the final reaction mixture. Each final reaction mixture contained 200 nM of enzyme, 500 μM of NADH, 500 μM of PEP, 1 mg/ml of bovine serum albumin (BSA), 10 mM of $MgCl_2$, 8 units of PK and 10 units of lactate LDH, 2 mM of I-1-P, and 20 mM Tris, pH 8.0. The total reaction mixture (50 μl) was incubated

for 2 minutes before adding UDP-GlcNAc. Ten measurements with respect to ten serial dilutions were taken at 340 nm (using a 96-well plate reader from Polarstar Omega, BMG labtech) for a period of thirty minutes. The linear portion of the kinetics was taken to calculate the reaction velocity, and it was plotted against donor substrate concentration at saturating acceptor concentration.

Acceptor kinetics was performed for both wild-type and mutant TbMshA, while holding the donor substrate (UDP-GlcNAc) at constant/saturating concentration (at 4 mM). The concentration of acceptor substrate (I-1-P) varied from 2 mM (final concentration) to 0 mM via serial dilution in the final reaction mixture. Each final reaction mixture contained 200 nM of enzyme, 500 μM of NADH, 500 μM of PEP, 1 mg/ml of bovine serum albumin (BSA), 10 mM of $MgCl_2$, 8 units of PK and 10 units of LDH, 4 mM of UDP-GlcNAc, and 20 mM Tris, pH 8.0. The total reaction mixture was incubated for 2 minutes before adding I-1-P. Ten measurements with respect to ten serial dilutions were taken at 340 nm (using a 96-well plate reader from Polarstar Omega, BMG labtech) for a period of thirty minutes. The linear portion of the kinetics was taken to calculate the reaction velocity, and it was plotted against acceptor substrate concentration at saturating donor concentration.

2.5 Mass spectrometry experiment for TbMshA

Mass spectrometry experiments for TbMshA and its mutant, E353Q, were done in the Laboratory for Biological Mass Spectrometry in the Department of Chemistry. Sample analytes from TbMshA and E353Q were extracted from *in vitro* reaction in

which 5 μl of enzyme (from stock concentration) was added into a 100 μl reaction buffer containing 20 mM Tris, pH 8.0, 5 mM (final concentration) of UDP-GlcNAc and I-1-P, incubated at room temperature for 2 hours before storing at 4^{o}C for overnight. Control sample analytes contained either the UDP-GlcNAc or I-1-P, dissolved in distilled water, with a final concentration of 50-60 μg/ml. All sample analytes were subjected to Electrospray Ionization (ESI) in both positive and negative modes, and were detected at the mass to charge (m/z) range between 100 to 700 (or 800). Due to the abundance of negative ions in the molecules (i.e. phosphate ions, hydroxyl ions), only the results from the negative ESI mode would be presented and discussed in this thesis.

3. RESULTS AND DISCUSSION

3.1 Determination of TbMshA structure via x-ray crystallography

In order to determine the crystal structure of MshA from *Mycobacterium tuberculosis* (herein termed as TbMshA), a protein crystal of TbMshA was obtained via vapor-diffusion method under silicon oil (which promoted the slow vaporization of the solvent) for 12 hours up to several days. A mountable size of the crystal (fitted in > 0.5 Å loop) was cryo-mounted (at 120 K) and shot by a monochromatic x-ray beam; the scattering of x-ray by the crystal (or diffraction) was recorded by the detector (image plate or CCD). The collected data was processed using macromolecular crystallography software HKL2000 (15). The whole data set was indexed into the proper space group, integrated and scaled accordingly to generate the reflection file ready for phasing.

Solving the crystal structure (or generating the electron density map which in turn is the Fourier Transform (FT) of the structure factor) requires both the amplitude and phase information from the diffraction data, but the phase of the diffraction data cannot be measured experimentally; hence there are computational techniques to determine the phase. One of the techniques to acquire phase information is known as molecular replacement (MR) (16).

MR is commonly used to calculate the initial phase from a structurally similar model (or known structure) in order to reconstruct the initial electron density for an unknown structure. For MR, a similar structure (typically > 30% identity is required) is used as a search model to perform three-dimensional rotation function and subsequent

three-dimensional translation function in the asymmetric unit cell to find the best fit solution. The rotation function uses Patterson function calculated from the search model and the observed data. This approach is used in the Molrep of CCP4 (17,18), a software suite used here to find the MR solutions for the apo and ligand-bound TbMshA structures.

For TbMshA apo structure, the phase was obtained by taking the C-terminal domain of ligand-bound TbMshA as an initial search model; once the best orientation was found, the C-terminal domain was fixed in the model and the N-terminal domain was again used as a secondary search model to obtain the overall solution for the entire monomer. For ligand-bound TbMshA, the phase was obtained directly by using the ligand-bound CgMshA (PDB code: 3c4v) as a search model.

3.2 Structural characteristics of apo TbMshA

The crystal structures of apo TbMshA is solved via x-ray crystallography to resolutions of 2.32 Å (Table 1) with a space group of $P2_12_12$, and unit cell dimensions of a = 159.3 Å, b = 56.6 Å, c = 106.6 Å, $\alpha = \beta = \gamma = 90^o$. The phasing of apo TbMshA was solved by molecular replacement (MR) with the ligand-bound TbMshA structure using CCP4, a collaborative software for macromolecular x-ray crystallography. There are two monomers (polypeptide chains) in the asymmetric unit of apo TbMshA. The monomer in the tertiary structure displays a typical β/α/β Rossmann fold (GT-B fold) seen in other members of glycosyltransferase superfamily (19). There are two distinct domains observed in the tertiary structure in which the N-terminal domain (or N-

domain) is having eight β-strands flanked by six α-helices, whereas the C-terminal domain (or C-domain) is having six parallel β-strands flanked by eight α-helices (Figure 5).

Table 1. Data collection and refinement statistics for apo and ligand-bound TbMshA.

Data collection	**TbMshA**	
Data set	**Apo**	**Ligand-bound**
Space group	$P2_12_12$	$P4_12_12$
Unit cell dimension (Å)	a=159.3 b=56.5 c=106.6 α=β=γ=90^o	a=123.6 b=123.6 c=154 α=β=γ=90^o
Completeness (%)	99.16 (39.83 - 2.32)	99.79 (30.0 - 2.89)
$I/I\sigma$	32.4	22.3
Redundancy	7.3	7.6
Refinement Statistics		
Resolution (Å)	2.32	2.89
R_{factor} (%)	21.3	20.4
R_{free} (%)	25.3	26
Small molecules	2 PO_4	2 UDP, 2 GlcNAc, 2 PO_4
Number of atoms	5999	6114
Number of reflections (total)	41911	27149
Unit cell volume (Å^3)	9.61 * 10^5	2.35 * 10^6
Root mean square deviations (rmsd)		
Bonds (Å)	0.005	0.005
Angles (^{o})	0.757	1.028

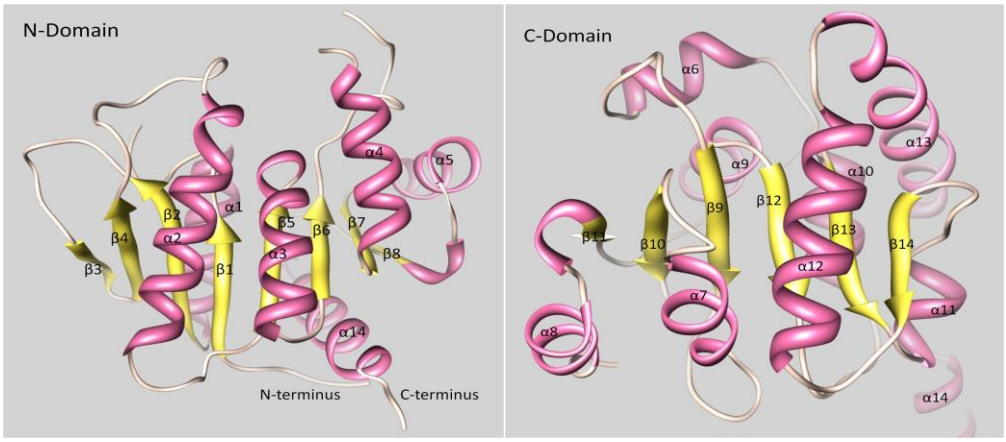


Figure 5. Domains of the apo TbMshA. Pink: alpha helix; yellow: beta strand; ivory: turns and loops.

The N-domain contains a mixture of parallel and anti-parallel β-strands, whereas only parallel strands are observed in the C-domain. A stretch of flexible loop is supposedly connecting the two domains, however, this region is highly flexible (believed to be the hot spot for ligand interaction), therefore the electron density is absent for this region in the apo structure.

3.3 Structural characteristics of ligand-bound TbMshA

The crystal structure of ligand-bound TbMshA is solved via x-ray crystallography to resolution of 2.89 Å (Table 1) with a space group of $P4_12_12$ and unit cell dimensions of a = 123.6 Å, b = 123.6 Å, c = 154.0 Å, $\alpha = \beta = \gamma = 90^{o}$. The phasing of ligand-bound TbMshA is solved using MR with its close homologue (with a sequence identity of 45%) in *Corynebacterium glutamicum*, CgMshA (PDB code: 3c4v). The asymmetric unit cell of ligand-bound TbMshA contains two polypeptide chains, with

each chain in its tertiary structure is displaying the same β/α/β Rossmann fold observed in the apo TbMshA. The N/C domains in ligand-bound TbMshA also comprise the same arrangement of β-strands flanked by α-helices as observed in the apo form (Figure 5). In addition, the loops that are invisible in the apo structure are become visible in the ligand-bound structure, Figure 6, indicated by the two arrows.

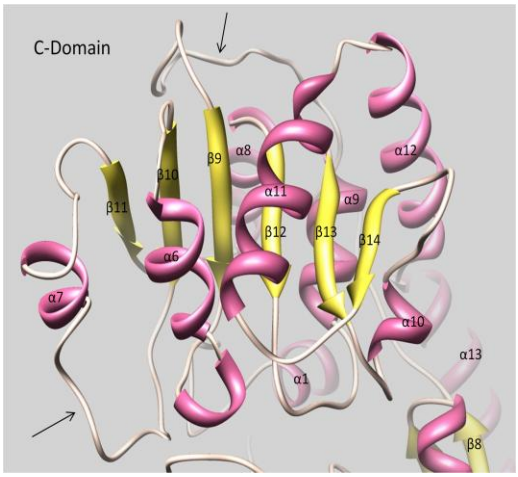


Figure 6. The C-terminal domain of the UDP-GlcNAc-bound TbMshA. Pink: alpha helix; yellow: beta strand; ivory: turns and loops.

The residues in the ligand-bound TbMshA structure that are interacting with the UDP-GlcNAc are Q331, H333, G67, L357, V358, E361, R273, K278, H178, E353, F355 and S354 (Figure 7). Among these, E353 is proposed here to be critical for the catalysis of TbMshA, as it is in close proximity of the N-acetyl-glucosamine moiety (believed to be transferred onto the acceptor substrate).

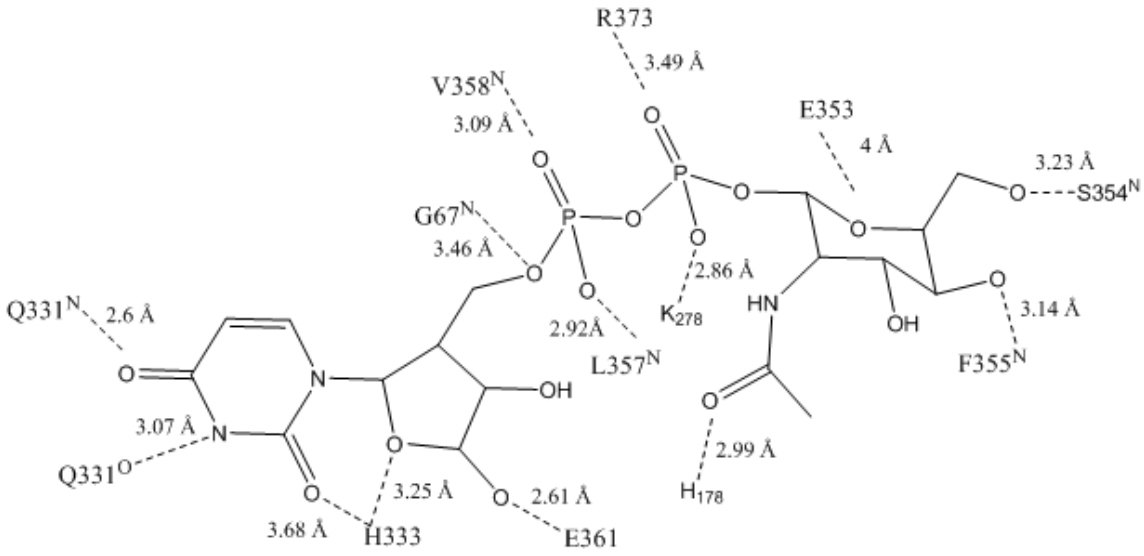


Figure 7. A schematic drawing of UDP-GlcNAc interacting with surrounding residues. Figure is prepared using ChemBioDraw Ultra 11.0.

3.4 Ligand-bound TbMshA reveals a C-domain movement

The asymmetric unit reveals the apo and ligand-bound TbMshA are two monomers in the crystallographic packing, and gel filtration chromatography (data not shown) confirms that TbMshA is monomeric in solution and is the functionally active form (20). Although the apo and ligand-bound TbMshA exhibit similar overall structure, the interface between the two monomers and the interdomain gap in the apo form is more open than the ligand-bound form. In addition, the loop region in which it connects the N/C domains is missing in the apo structure (Figure 8).

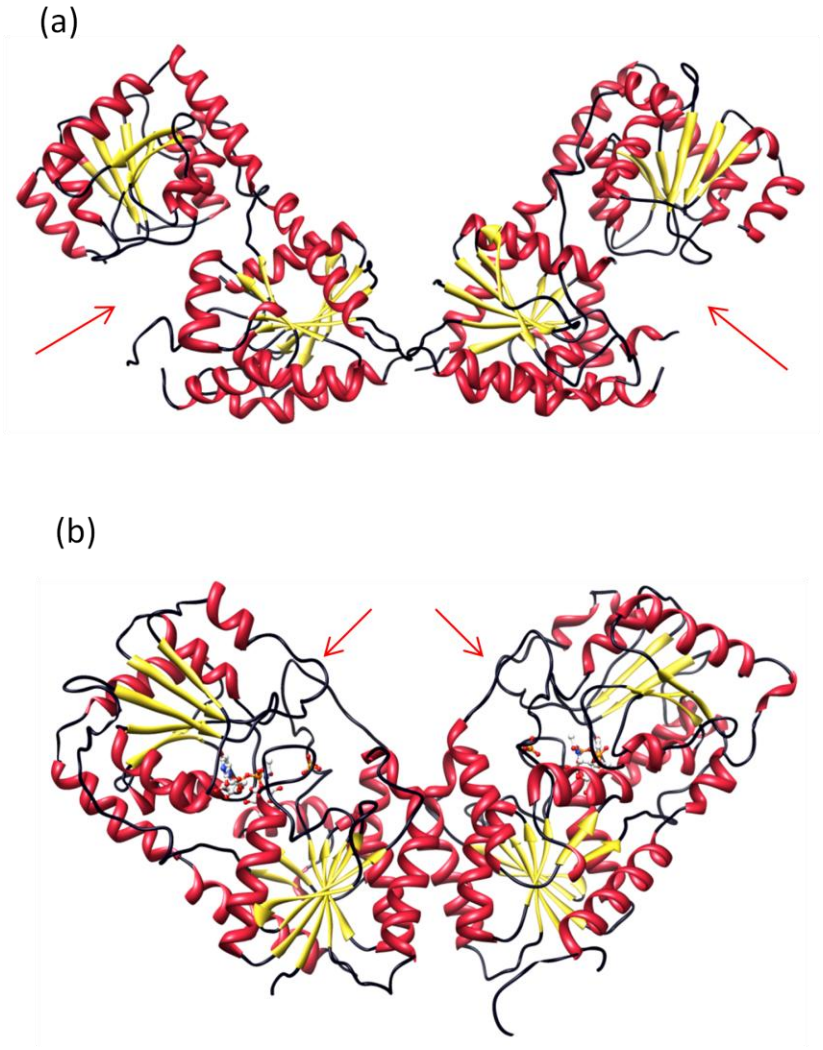


Figure 8. Overall structure of apo (a) and ligand-bound (b) TbMshA in ribbons. Red: alpha helices; yellow: beta strands; black: loops and turns. Red arrows indicate the loop region that is missing in the apo TbMshA structure.

The two domains in the closed from are facilitated by several highly flexible loops that are not properly shown in the absence of ligand (and thus lack electron densities for model building, Figure 8a). The flexible loops are constituted by domain

interface residues such as, Pro-60 to Gly-66, Asn-185 to Glu-195, and Gly-299 to Ala-308. These loops are being stabilized upon UDP-GlcNAc binding, where the ligand is situated in a cleft caged by the mobile loops that are not shown in the apo structure (Figure 9a). A 90^{o} rotation along either the horizontal or vertical axis (Figure 9b and 9c) shows the observable secondary structures movement upon ligand binding in the C-domain, especially α7 and α12 which become α6 and α11 (see the figures on p. 15-16).

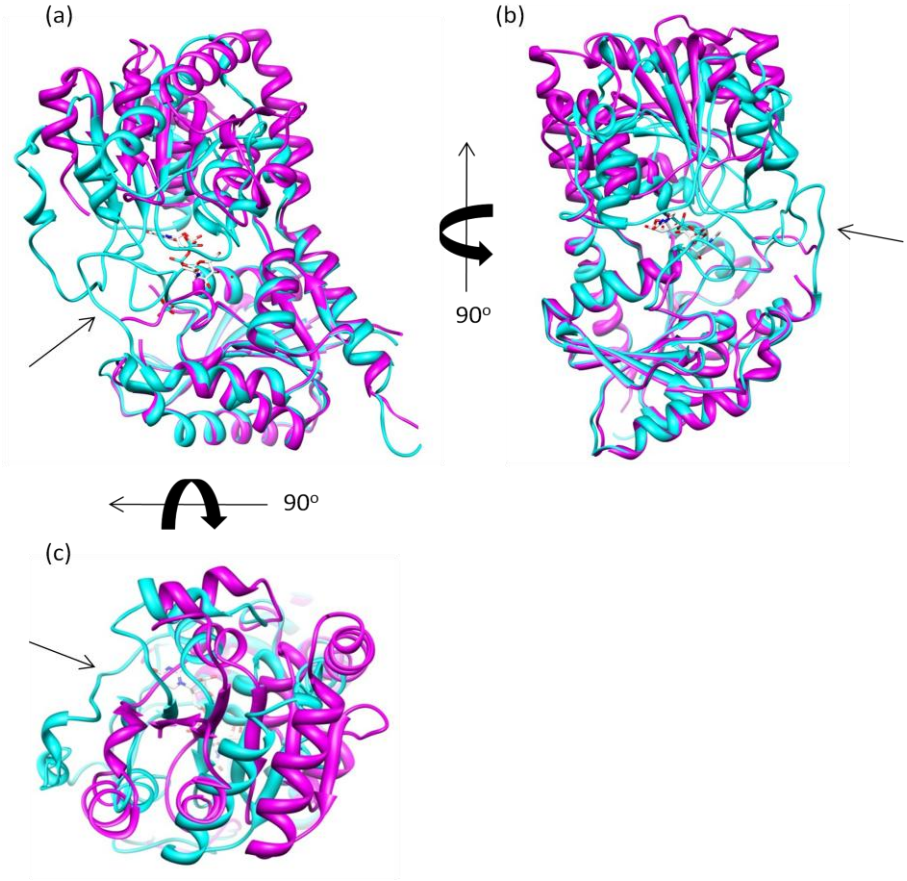


Figure 9. Comparison of the monomeric structure of apo and ligand-bound TbMshA upon vertical (b) and horizontal (c) rotation. Magenta: apo TbMshA; aqua: ligand-bound TbMshA.

The N-domain of TbMshA is relatively rigid and conserved structurally, as compared with its structurally known homologue, CgMshA (Figure 10).

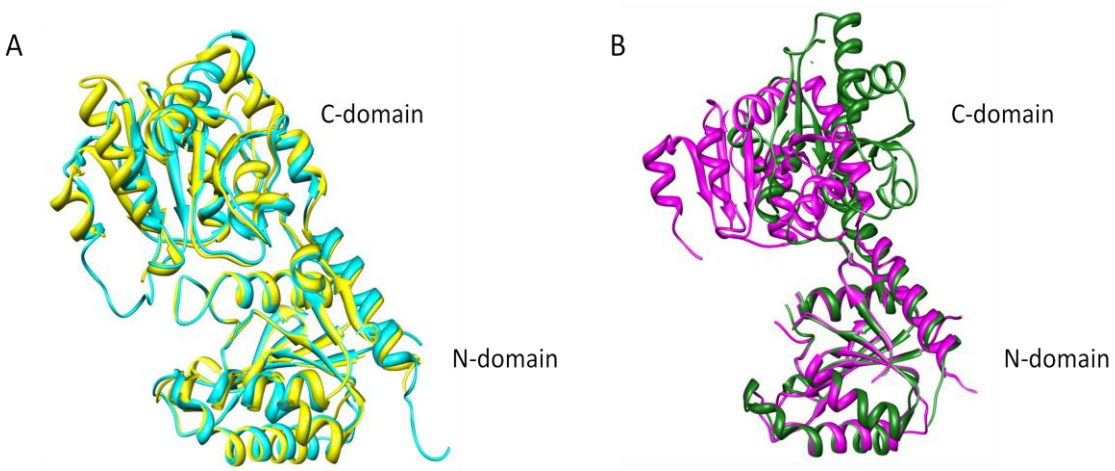


Figure 10. Superimposition of TbMshA with CgMshA. A. ligand-bound TbMshA (aqua) superimpose with ligand-bound CgMshA (yellow, PDB code: 3c4v). B. apo TbMshA (magenta) superimpose with apo CgMshA (green, PDB code: 3c48).

The C-terminal domain, however, undergoes some structural movement upon UDP-GlcNAc binding albeit such conformational shift is less drastic as the one seen in the case of CgMshA (Figure 10B). This conformational shift is manifested via the obvious distortion of two helices (α6 and α8) in the apo TbMshA (see the figures on p. 15-16). The transformed loop from helix 8 becomes parts of the mobile loop region that serve as a closing lid to encapsulate the bound UDP-GlcNAc (Figure 9). The nearby loops are not able to be visualized by electron density, rendering the inaccessibly to determine the extent of loop mobility. However, such conformational change (especially

hinge movement/loop closing) upon ligand and/or substrate binding is a characteristic feature observed in members of glycosyltransferase superfamily (21).

The observation of C-terminal domain movement upon ligand bound in TbMshA is not purely a crystallographic artifact due to the choice of reference. The superimposition of the N- and C-terminal domain shows non-overlapping margins (see Appendix C on p. 54) even though both domains display similar β/α/β fold, indicating the shift of the secondary structure (helices) is domain specific upon ligand binds. The remarkable differences are unquestionably observed in the loop regions.

3.5 UDP-GlcNAc is hydrolyzed in the ligand-bound TbMshA crystal structure

Previous refinement assumed that the UDP-GlcNAc was intact in the ligand-bound TbMshA structure (see Appendix A on p. 51), however, after rounds of new refinement, the electron density of the ligand-bound TbMshA supports the "hydrolysis" of UDP-GlcNAc in which the GlcNAc moiety is separate from the UDP-GlcNAc (Figure 11).

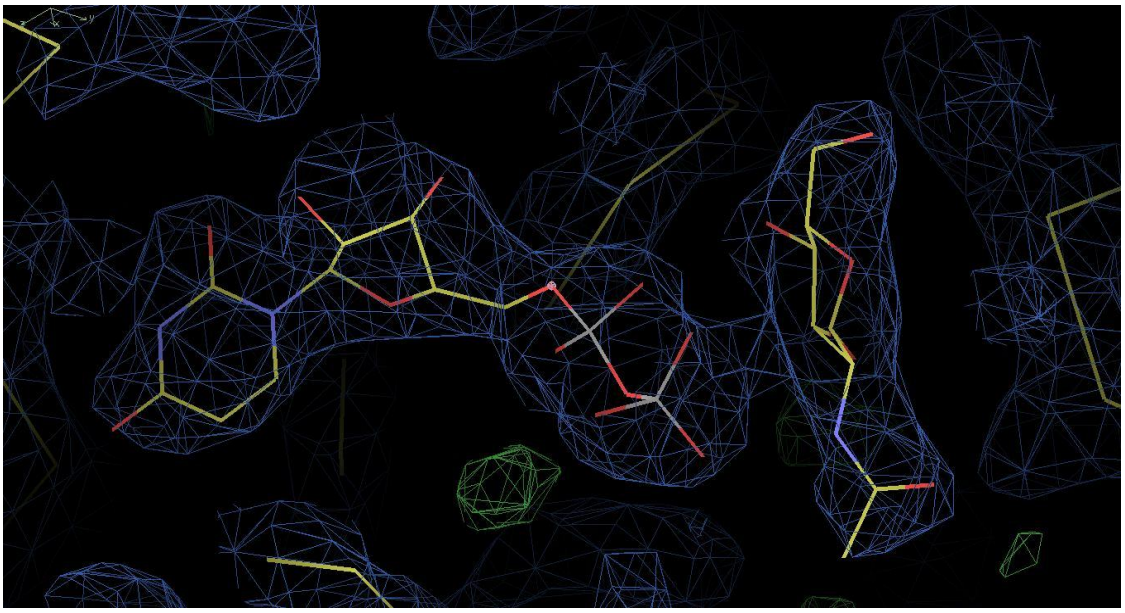


Figure 11. The electron density map of the ligand UDP-GlcNAc shown as two separate pieces. The density map is a Fo-Fc map generated from COOT, contoured at 1.7 σ. Left: UDP, right: GlcNAc.

The original proposed ligand-bound structure of TbMshA was with an intact UDP-GlcNAc, however, after careful inspection and many rounds of refinement, the intact piece was unsupported by the new refinement as the linkage between the β phosphate of UDP and the GlcNAc moiety was always appeared as negative density (from the Fo-Fc map, data not shown). So, breaking the linkage in between the UDP-GlcNAc in new refinement runs appeared to be plausible as the negative density was gone and the resulted in a decrease of the R factor (a parameter that compares the overall agreement between the calculated model and the actual observation). The whole ligand was refined as two separate pieces, after which the GlcNAc moiety seemed to undergo a

flipping motion. This flipping renders the GlcNAc adopts the most stable confirmation (chair confirmation) and is still surrounded by the same residues as depicted by the figure on p. 17. The observation of this cleavage doesn't seems to be arbitrary or due to hydrolysis, because UDP-GlcNAc is still remained intact in water, as confirmed by Electrospray Ionization experiment (see the figure on p. 39). Only in the presence of enzyme and acceptor substrate would it seem to be behaving this way. This behavior may also suggest that TbMshA is undergoing a bi-bi ping pong mechanism in which the first by-product (UDP) leaves before the uracil moiety is transferred onto the acceptor substrate.

3.6 Ligand-bound TbMshA structure reveals probable active site

The interactions between UDP-GlcNAc and nearby residues in the loop region stabilize the mobility in which it is not observed in the apo form. Such ligand-bound stabilization is also a representative phenomenon exhibited by other glycosyltransferase members (19). This ligand-bound stabilization somehow enables the identification of several critical residues involved in the probable active site cavity (see the figure on p. 17). The uracil moiety forms polar interactions with the amide and carbonyl backbone of Gln-331, as well as the side chain of Asn-59. The imidazole ring of His-333 helps to mediate the carbonyl oxygen and the ribose oxygen. The 2', 3'-hydroxyl group on the ribose are interacting with Glu-361, possibly through hydrogen bonding. The diphosphate groups are stabilizing by the charged residues, Arg-273, Lys-278 at the beginning of α6 and the amide nitrogen from the backbone of Gly-67, Leu-357 and Val-

358 (located at the domain interface) respectively. The refined electron density supports the cleavage of GlcNAc from UDP, but it does not alter the surrounding environment of the N-acetyl-glucosamine moiety. In fact, this refinement allows the ribose ring to adopt the most stable confirmation (chair confirmation), and it is now flipped over, having its pyran oxygen facing away from the β phosphate of the UDP such that the pyran oxygen is face-to-face with the δ nitrogen of the His-178, while the acetyl amide tail is interacting with Glu-353 and the amide backbone of Ser-354, and a phosphate group (from soaking) is at close proximity (a distance of 4.0 Å) from the 2'-hydroxyl group (Figure 12).

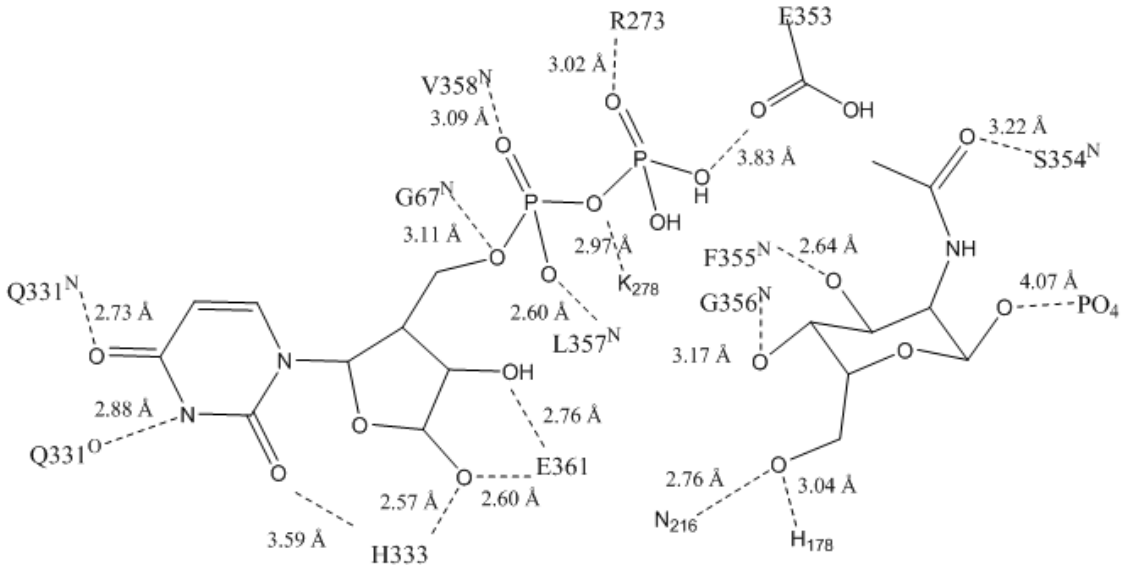


Figure 12. A schematic drawing of UDP-GlcNAc as two separate pieces with the surrounding residues based on the ligand-bound TbMshA structure. Figure is prepared using ChemBioDraw Ultra 11.0.

The 3’ and 4’-OH of GlcNAc are facing the amide nitrogen of Phe-355 and Gly-356, interacting thru hydrogen bonding (the distances between the amide backbone to the hydroxyl groups are 2.64 Å and 3.17 Å respectively). The 6’-OH is most likely interacting with the δ nitrogen of His-178 through hydrogen bonding. At a distance of about 4 Å from the 1’-OH of the GlcNAc moiety, a phosphate (from soaking) is fitted into the density in the ligand-bound TbMshA structure. This phosphate group is surrounded and stabilized by a group of charged residues, such as Lys-184, Lys-122, Thr-179, Arg-199 and Tyr-155 that form a small binding cavity in the vicinity of the UDP-GlcNAc binding pocket; it seems to have a long-range interaction with the 1’-hydroxyl group of GlcNAc (Figure 13).

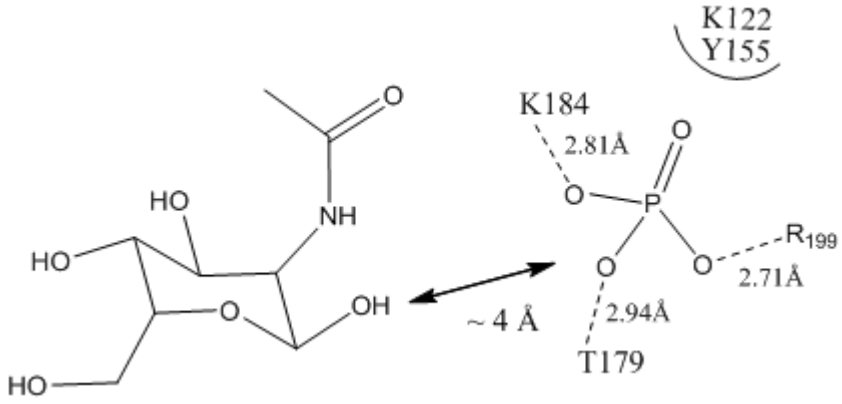


Figure 13. A schematic drawing of the phosphate “pocket” with its surrounding charged residues, at a distance that is around 4 Å from the GlcNAc moiety. Figure is prepared using ChemBioDraw Ultra 11.0.

Because the phosphate is about 4 Å away from the pyran oxygen of the GlcNAc moiety, its occupancy in the cavity may have coincided with the 1'-phosphatyl group of I-1-P (see Appendix A on p. 51). This configuration would have been enabled the sugar moiety of I-1-P in a direct face-to-face interaction with the sugar moiety of the UDP-GlcNAc, whose sugar moiety has been considered as the responsible sugar donor that is transferred onto the substrate (20). However, because there is no crystallographic evidence to show where I-1-P binds in the ligand-bound TbMshA structure (due to failure in obtaining crystals in this regard), the mode of action for the GlcNAc transfer remains speculative for TbMshA at this point.

3.7 E353 may be the nucleophile for catalysis in TbMshA

Based on the crystal structure of the ligand-bound TbMshA, Glu-353 is in the proximity of the β phosphate of UDP and the pyran oxygen of the GlcNAc moiety (see the figure on p. 24). More importantly, this residue can serve as a nucleophile in the protein crystal (given the pKa for glutamate side chain is 4.1 and the pH of the crystal condition is 8.5). In the previous refined structure (data not shown), there was an ordered water molecule (within hydrogen bonding distance) in the proximity of Glu-353. Therefore, it was proposed to increase the nucleophilicity of a water to attack the C1' position of the pyran. However, in the re-refined structure of ligand-bound TbMshA, this ordered water is not seen anymore and the UDP-GlcNAc is shown as pieces. One may speculate that the splitting of UDP-GlcNAc may due to its high hydrolytic profile in solution. But, mass spectrometry indicates that UDP by itself is not observed in the

spectrum in aqueous solution, indicating that UDP-GlcNAc is not hydrolyzed easily in solution (see the figure on p. 39). But interestingly enough, UDP-GlcNAc is not hydrolyzed either into UDP and GlcNAc in only the presence of the enzyme without the acceptor substrate (see the figure on p. 41). Hence, even if one wishes to propose the mechanism is bi-bi ping pong, the intermediate product is not readily released until the acceptor substrate has come in. However, in this case, phosphate was used during crystallization to mimic the presence of the I-1-P, then the split UDP-GlcNAc was observed. Although the attempt to co-crystallize with both substrates (UDP-GlcNAc and I-1-P) had failed to produce diffraction-quality crystals for data collection, based on the tertiary complex obtained from the pieces of UDP-GlcNAc and phosphate, as well as the known structural information from a close homologue (CgMshA), a mechanism is proposed (Figure 14). Glu-353 located around 4 Å from the anomeric carbon of the GlcNAc moiety is believed to be a nucleophile to initiate the attack, making the electrons readily withdrawn to the oxygen in the β phosphate of UDP, while the diphosphate groups are charged stabilized by the nearby basic amino acids (R273 and K278). The proposed intermediate state may involve a covalently bound N-acetyl-glycosyl-enzyme intermediate to stabilize the transition state before the final attack from the 3'-OH of I-1-P to form the product (GlcNAc-Ins-P). Since there is no direct crystallographic evidence to support this mechanism, a composite figure (see Appendix A on p. 51) from ligand-bound TbMshA and CgMshA structure is generated using Chimera to illustrate the interactions that are consistent with the scheme shown in Figure 14.

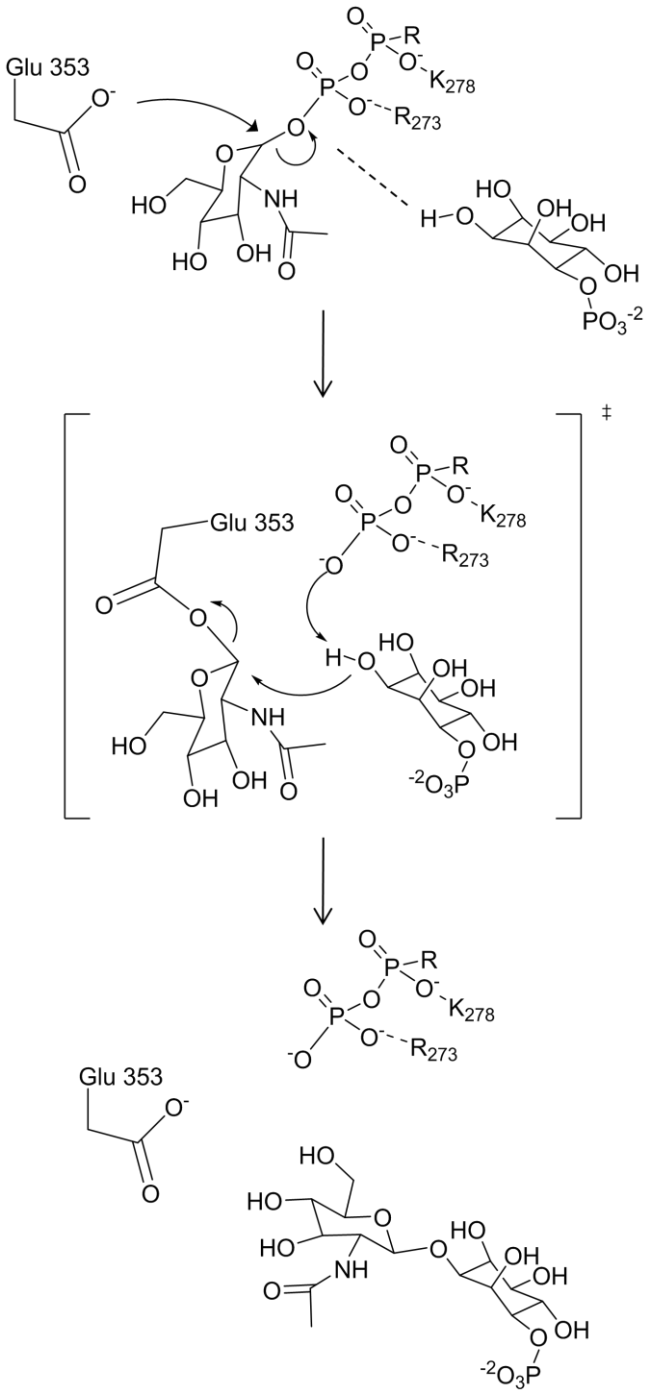


Figure 14. Proposed mechanism for the transfer of the glucosamine moiety onto the substrate I-1-P. The uracil moiety from UDP is represented as R. Figure is prepared using ChemBioDraw Ultra 11.0.

Based on the current refined ligand-bound TbMshA structure, the diphosphate group is stabilized by flanking charged residues (Lys-278, Arg-273, and the amide backbone of Gly-67, Leu-357 and Val-358), the diphosphate group is therefore ready to withdraw electrons, facilitating the formation of the transition state.

The TbMshA crystal structure solved with UDP-GlcNAc cannot fully depict TbMshA's sugar donor-to-acceptor mechanism, partly because the structure is solved at a relatively low resolution (2.89 Å). Also, even though a pseudo substrate, myo-inositol, was used for soaking, no extra positive density that corresponded to the size of the molecule was found. This may infer that myo-inositol is not a good substrate analogue for the acceptor. Although the exact details on the transferring mechanism occurred for TbMshA is entirely speculative at this point, a mutation purposely introduced for TbMshA at residue 353 renders the mutant form (E353Q) hinders its enzyme activity compared to the wild-type as well as loosing the capability to produce the product (see Section 3.9-3.11).

3.8 TbMshA vs. CgMshA

Through sequence homology search, CgMshA, the glycosyltransferase from *Corynebacterium glutamicum* displays the greatest extent of sequence identity (~45%, Figure 15) and structure similarity (especially in the ligand-bound structure, see the figure on p. 20).

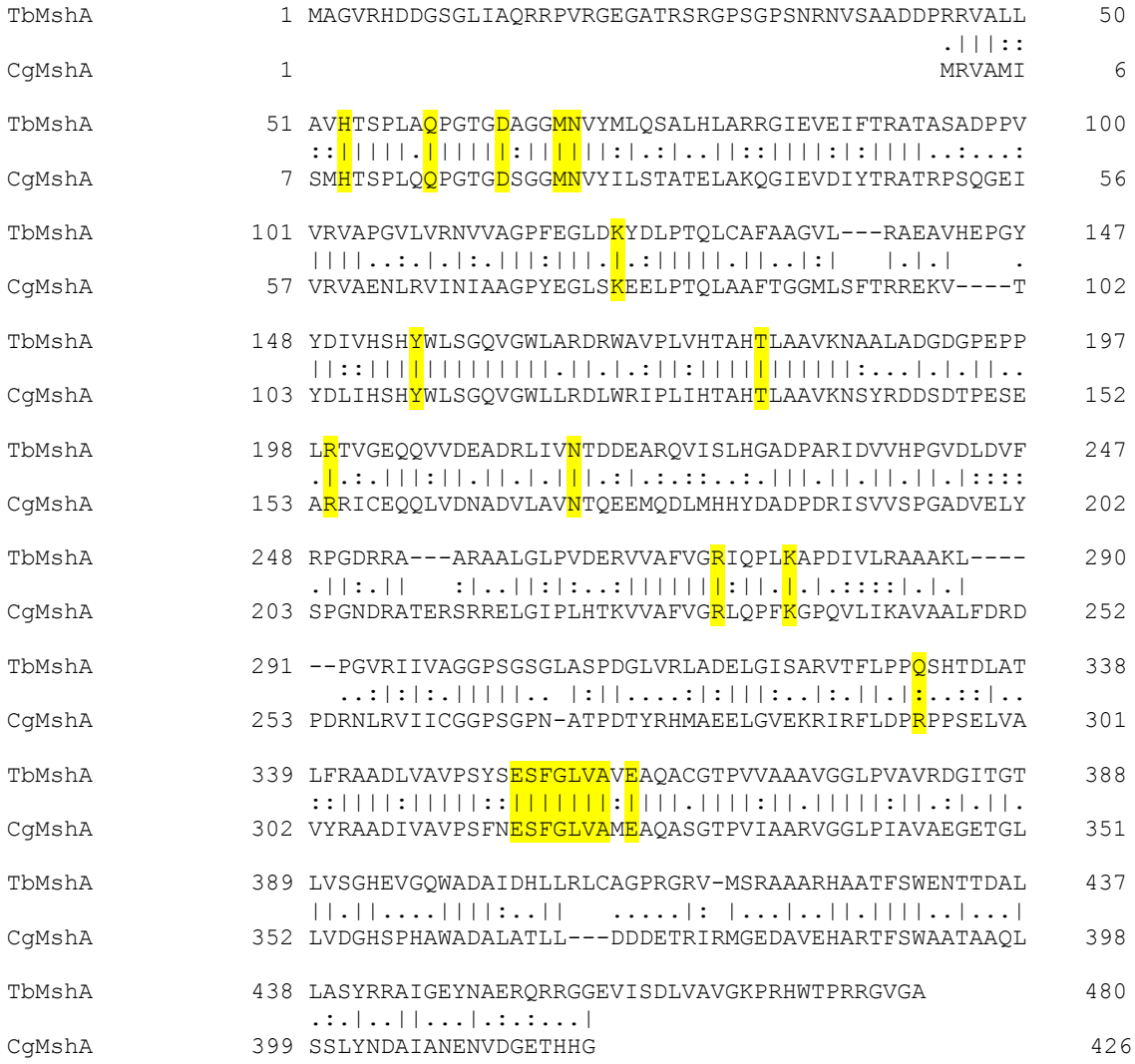


Figure 15. Sequence alignment of TbMshA and CgMshA. The conserved residues responsible for interacting with the ligand (UDP-GlcNAc) are highlighted. "|" indicates identical residues.

In the apo form of CgMshA (PDB code: 3c48), the C-terminal domain undergoes a substantial conformational change (> 90^{o} rotation) with respect to the N-terminal domain. The apo CgMshA is also more open compared to the ligand-bound CgMshA

(22). The ligand-bound structure (PDB code: 3c4v) contains the electron density for UDP and the acceptor substrate I-1-P but not the GlcNAc moiety, which is later modeled in the ligand-bound structure. Residues that are interacting with the UDP-GlcNAc in the CgMshA model are conserved in the TbMshA (Figure 15), except Arg-294 in CgMshA where it is being replaced by Gln-331 in TbMshA, which has hydrogen bond interaction with the 2' amine in the uracil moiety. Because the full length molecule of UDP-GlcNAc is modeled in and it is not derived from the electron density in the ligand-bound CgMshA structure, the actual orientation of the GlcNAc moiety with respect to the UDP is also speculative in the case of the ligand-bound CgMshA structure. In the ligand-bound TbMshA structure, although the electron density for the whole molecule seems to be connected, the refinement result supports the hydrolysis of the bond between the UDP and GlcNAc moiety. Hence, with this hydrolysis, the orientation of the GlcNAc is 180^{o} flipped compared to the original proposed structure (see the figures on p. 17, p. 24 and Appendix A on p. 51). In addition, this flipping over of the GlcNAc makes it possible for the 1' hydroxyl group facing the phosphate pocket in a near distance (see the figure on p. 25). Although still lacking substantial crystallographic evidence, this orientation in fact is in favor of the proposed mechanism (see the figure on p. 28). Furthermore, a cavity suggested by CgMshA which consisted of a cluster of charged and polar residues for the acceptor substrate is also coincidentally occupied by the phosphate group in the TbMshA ligand-bound structure. Whether this is actually the site that will be occupied by the substrate, I-1-P, is not definite (only a composite figure is generated to mimic the situation in crystal structure of TbMshA). However, should the phosphate group from

the I-1-P occupy the same phosphate pocket as observed in the TbMshA ligand-bound structure, UDP should be leaving the cavity first, allowing more rooms for the other residues to move around, especially allowing the GlcNAc moiety to be pushed forward and be posed itself against the I-1-P for bonding. If this is the case, such mechanism is not depending on a S_Ni-substrate assisted mechanism which involves an oxocarbenium-ion like transition state, as observed in the case of CgMshA (22), but rather, it follows through a more classical retaining glycosyltransferase mechanism in substrate transfer as seen in the cases of other retaining glycosyltransferases (19,23).

3.9 TbMshA enzyme assay

To date, there are no kinetic parameters obtained directly for TbMshA, except for that of the closest homologue of MshA in *Corynebacterium glutamicum* (CgMshA) provided by recent literature (22). Hence, this part of the thesis attempts to obtain the kinetic parameters, such as K_m, V_{max}, k_{cat} and k_{cat}/K_m for TbMshA. In addition, the finding of kinetic parameters would facilitate the comparison of TbMshA with its homologues in other *Mycobacterium* species, as well as to provide the preliminary data needed to investigate the activities of TbMshA and study its inhibition in the near future.

Since TbMshA belongs to the glycosyltransferase superfamily, substrates of these superfamily members are usually donor-acceptor pairs, in which the donor is generally nucleotide-based sugar, and the acceptor is variable upon the catalytic specificity of each member. TbMshA utilizes the UDP-GlcNAc – I-1-P donor-acceptor pairs, in which the GlcNAc moiety will be first hydrolyzed by the enzyme, generating

free UDP. The GlcNAc moiety will then be transferred onto the acceptor, I-1-P, generating the first intermediate product, GlcNAc-inositol-phosphate (GlcNAc-Ins-P), which is the first intermediate in the mycothiol biosynthesis pathway. The free UDP generated from TbMshA then becomes the substrate of another enzymes pair: pyruvate kinase (PK) and lactate dehydrogenase (LDH), whose reaction would oxidize the reducing agent NADH to NAD^+; hence, the TbMshA assay is performed via a coupled enzyme assay fashion (23). The spectroscopic measurement of the enzyme activity is feasible via the oxidation of NADH to NAD^+ at 340 nm, as shown in Figure 16:

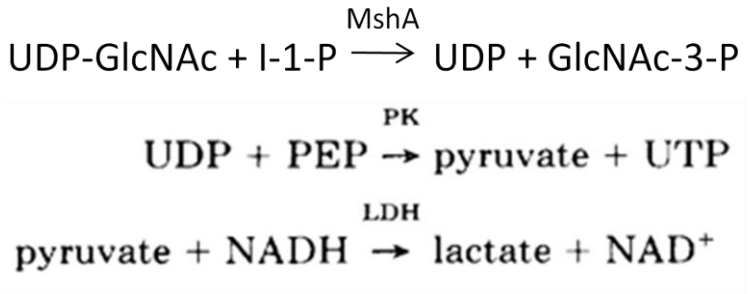


Figure 16. A schematic representation of the coupled-enzyme reaction carried out by TbMshA, at the expense of oxidizing NADH to NAD^+.

In order to determine the basic kinetic parameter for TbMshA, the enzyme reaction rate with respect to the substrate concentration is measured. Since TbMshA utilizes two substrates, the kinetic profile is plotted for one substrate while holding the other constant at saturation. Each profile demonstrates the Michaelis-Menten kinetics assuming steady state (change of enzyme-substrate complex remains constant over time). The Michaelis-Menten equation (Equation 1) models the kinetics such that v is

the reaction rate (in μM sec^{-1}) at any given substrate concentration [*S*] (in μM), V_{max} is the maximum velocity (in μM sec^{-1}), K_m is the Michaelis-Menten constant (in μM), which corresponds to the concentration of substrate at half V_{max}.

$$v = (V_{max}\,[S])/(K_m + [S]) \qquad \text{(Eq.1)}$$

In order to plot the reaction rate in the proper unit (μM/sec), the raw data plot, which measures the enzyme activity by obtaining the slope of decline in absorbance at 340 nm over time (30 minutes), is converted into the rate at which the product is formed. Based on Figure 16 (on p. 33), the formation of NAD+ is in 1:1 molar ratio with UDP, which is generated from TbMshA, hence, proper [NAD+] calibration (data not shown) allows one to convert the spectroscopic slope into reaction rate in micro-molar concentration per unit time. Another way to figure out the K_m and V_{max} of the enzyme kinetics is to take the reciprocal of the Michaelis-Menten equation (Equation 2):

$$1/v = (K_m/V_{max})\,(1/[S]) + 1/V_{max} \qquad \text{(Eq.2)}$$

The reciprocal plot or Lineweaver-Burk plot (see Appendix B-1 and B-2) allows determination of the K_m and V_{max} directly from the slope and the y-intercept.

3.10 WT vs. mutant (E353Q) TbMshA

The coupled enzyme assay is also performed in the same manner for the mutant TbMshA, E353Q, and its kinetic profile (following Michaelis-Menten kinetics) is plotted against the wild-type (Figure 17 and Figure 18) for comparison.

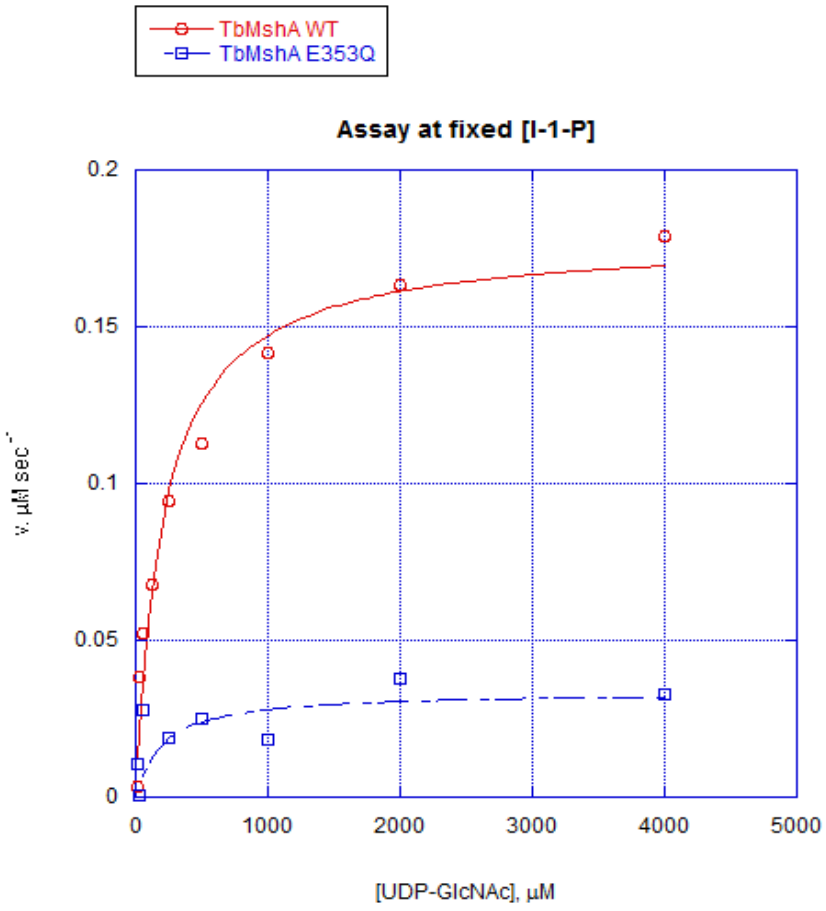


Figure 17. Plot of reaction rate (in µM sec^{-1}) versus UDP-GlcNAc concentration while holding I-1-P constant at 2 mM. Open circle: wild-type (WT) TbMshA; open square: mutant TbMshA, E353Q. Figure is generated using KaleidaGraph 4.0.

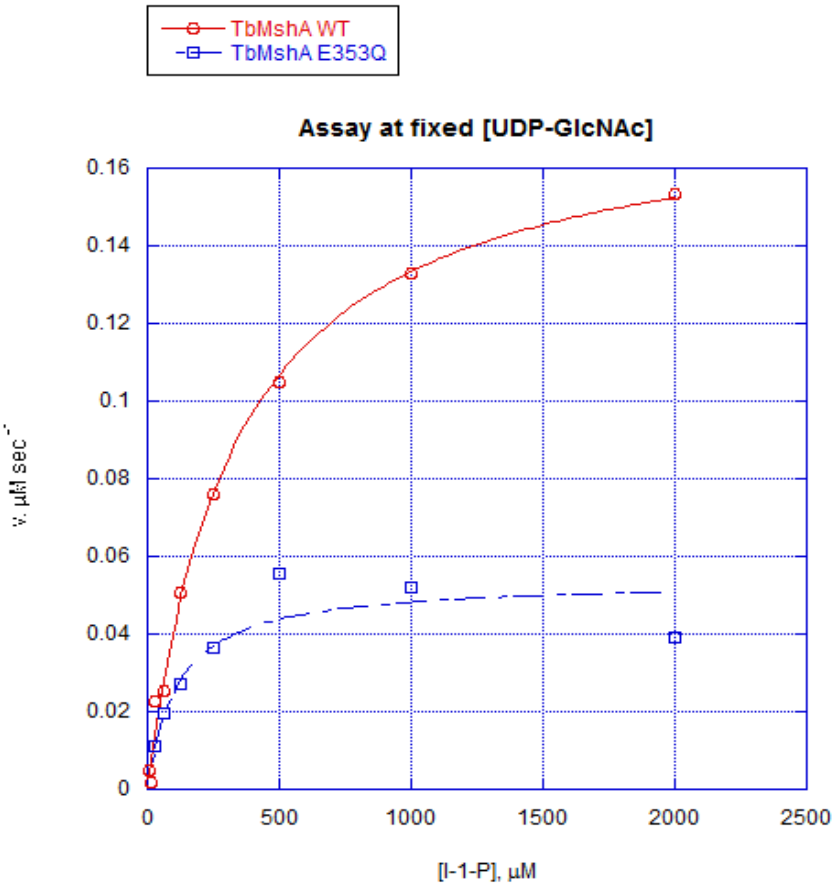


Figure 18. Plot of reaction rate (in µM sec^{-1}) versus I-1-P concentration while holding UDP-GlcNAc constant at 4 mM. Open circle: wild-type TbMshA; open square: mutant TbMshA, E353Q. Figure is generated using KaleidaGraph 4.0.

The activity of the mutant is significantly quenched compared to the wild-type. The apparent maximum velocity, V_{max}, is reduced by more than four-fold in the mutant form (Table 2). The values of K_m and V_{max} can be directly obtained from the Lineweaver-Burk plot, and k_{cat}, the turnover number of an enzyme, is calculated from V_{max} divided by the total concentration of enzyme, $[E]_T$, such that (Equation 3):

$k_{cat} = V_{max}/[E]_T \approx V_{max}/[E]$ (Eq.3)

The total enzyme concentration $[E]_T$ in the reaction is the sum of enzyme-substrate complex concentration [*ES*] and free enzyme concentration [*E*], but assuming steady state, [*ES*] does not change over time (because it returns to free enzyme as soon as product forms), then $[E]_T$ is approximately equal to [*E*], which is 0.2 μM in this case.

Based on the kinetics data, the Michaelis-Menten constants (K_m) for UDP-GlcNAc and I-1-P are found to be 210 μM and 331 μM respectively. The K_m found for the mutant, E353Q, is 204 μM for UDP-GlcNAc, 113 μM for I-1-P. The V_{max} for wild-type TbMshA is found to be 0.17 ± 0.02 μM sec^{-1}, which is four-fold higher than that for E353Q (0.04 ± 0.01 μM sec^{-1}). The turnover number, k_{cat}, is 0.9 ± 0.1 sec^{-1} for wild-type, and 0.2 ± 0.1 sec^{-1} for E353Q. k_{cat}/K_m, which provides an index of enzymatic catalytic efficiency, is calculated to be 3.5 ± 1.1 * 10^3 M^{-1} sec^{-1} for the wild-type TbMshA and 0.3 ± 0.02 * 10^3 M^{-1} sec^{-1} for the mutant. Table 2 summarizes the kinetic parameters found for the wild-type and mutant TbMshA:

Table 2. Kinetic parameters for WT and mutant (E353Q) TbMshA.

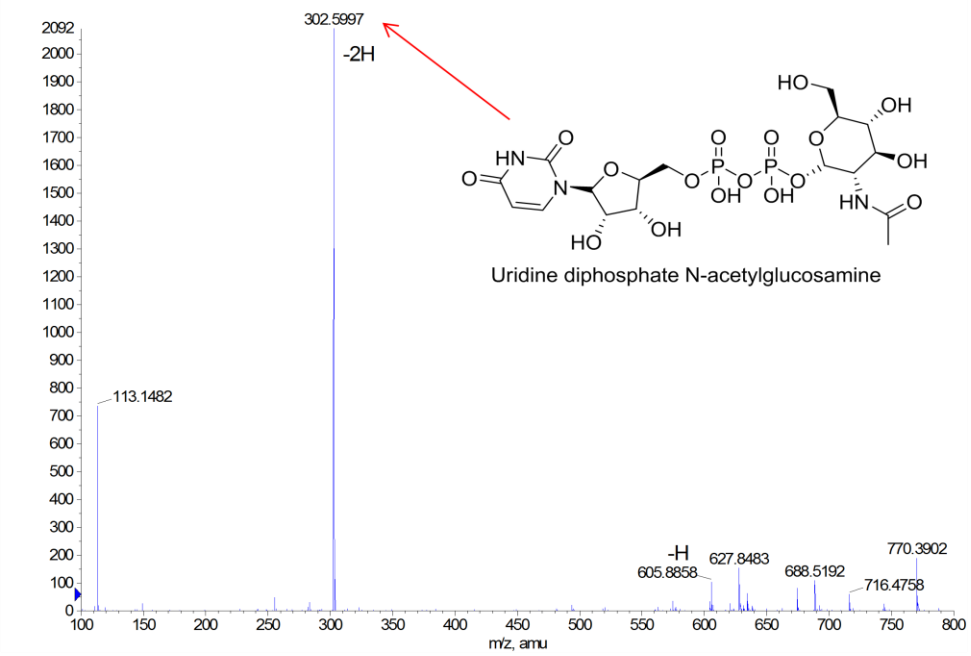
TbMshA	**K_m, μM**				
	UDP-GlcNAc	**I-1-P**	**V_{max}, μM sec^{-1}**	**k_{cat}, sec^{-1}**	**k_{cat}/K_m, M^{-1} sec^{-1}**
WT	210	331	0.17 ± 0.02	0.9 ± 0.1	3.5 ± 1.1 * 10^3
E353Q	204	113	0.04 ± 0.01	0.2 ± 0.1	0.3 ± 0.1 * 10^3

The kinetic parameters reveal that the mutant has displayed a greatly reduced enzyme activity. The V_{max} for E353Q are four-fold lower than that of the wild-type, and the k_{cat} of E353Q is even an order of magnitude lower compared to that of the wild-type. Although the K_m for UDP-GlcNAc for both enzymes is comparable, since the mutant has a greatly reduced k_{cat}, the catalytic efficiency (k_{cat}/K_m) of the mutant is still about four-fold lower than that of the wild-type, indicating that the mutation at residue 353 is indeed impairing the native enzyme activity.

3.11 Mass spectrometry analysis of TbMshA

Since the coupled enzyme assay could only detect the liberation of UDP to demonstrate the activity of TbMshA, there is actually no definite answer as to whether TbMshA would get to make the product in the reaction. In this regard, there is a need to find out whether both wild-type and mutant form of TbMshA can make product, even though the enzyme activity of the mutant is greatly hindered. *In vitro* reaction products of wild-type TbMshA and its mutant form, E353Q, are subject to Electrospray Ionization (ESI, negative mode) to trace the presence of the immediate product, GlcNAc-Ins-P. The results indicate that wild-type TbMshA is able to produce GlcNAc-Ins-P ($C_{14}H_{25}NO_{14}P$, with a negative charge, which has an m/z of 462). But this product is not observed in the case of the mutant, E353Q, since there is no corresponding peak at this m/z value. The controls used in this ESI experiment are the two substrates, UDP-GlcNAc ($C_{17}H_{27}N_3O_{17}P_2$, Mw = 607 g/mol, m/z = 302.5), and I-1-P ($C_6H_{11}O_9P_2$, Mw = 259 g/mol, m/z = 259.1) (Figure 19).

A)



B)

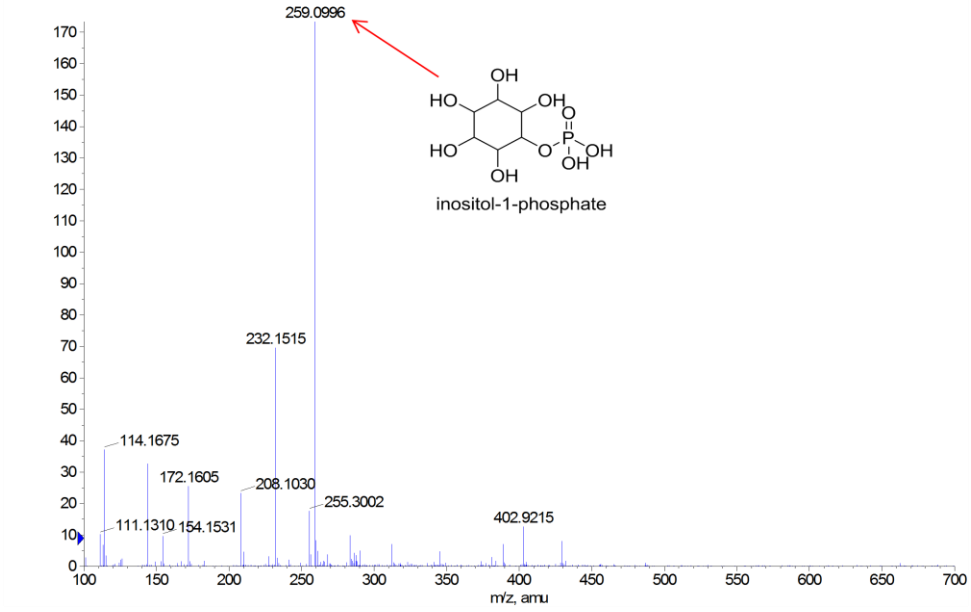
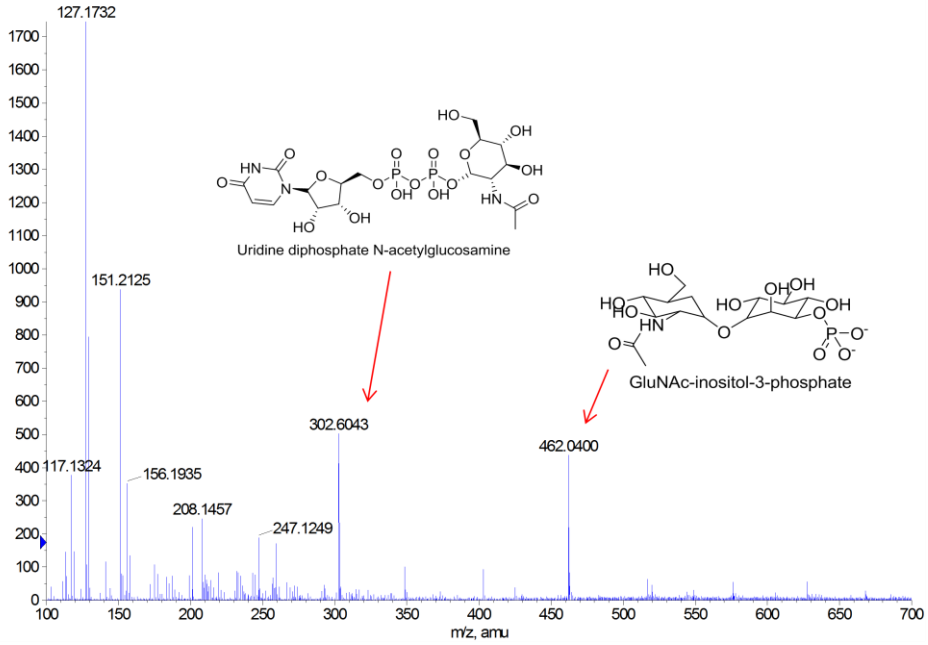


Figure 19. ESI spectra (negative mode) of A) UDP-GlcNAc, B) I-1-P, C) WT TbMshA, and D) mutant form of TbMshA, E353Q.

C)



D)

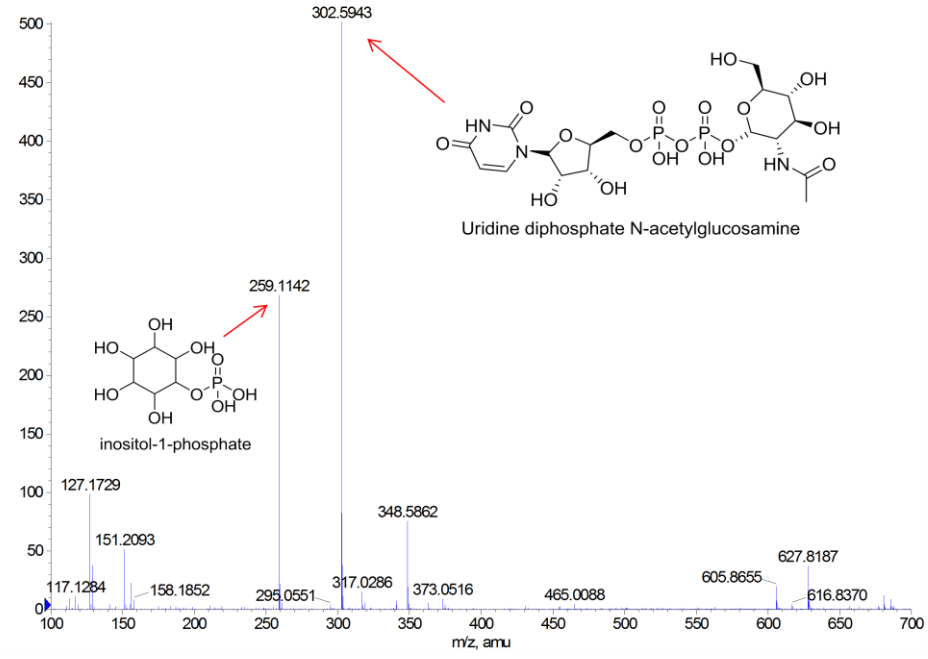


Figure 19 continued.

The wild-type TbMshA shows a peak corresponds to the mass of its immediate product, GlcNAc-Ins-P, at 462 m/z (assuming a unit charge), while in the ESI profile of the mutant form, E353Q, this peak is absent, indicating that the E353Q cannot turn over the substrates. In the kinetic assay, TbMshA is believed to first release the UDP from UDP-GlcNAc before the GlcNAc moiety is transferred onto I-1-P. This "release first" mechanism is supported by the crystal structure of ligand-bound TbMshA, however, this hypothesis is not fully supported because the ESI results still show an "intact" molecule of UDP-GlcNAc (no peak corresponds to free UDP) when the reaction mixture only contains UDP-GlcNAc and enzyme without the acceptor substrate, I-1-P (Figure 20).

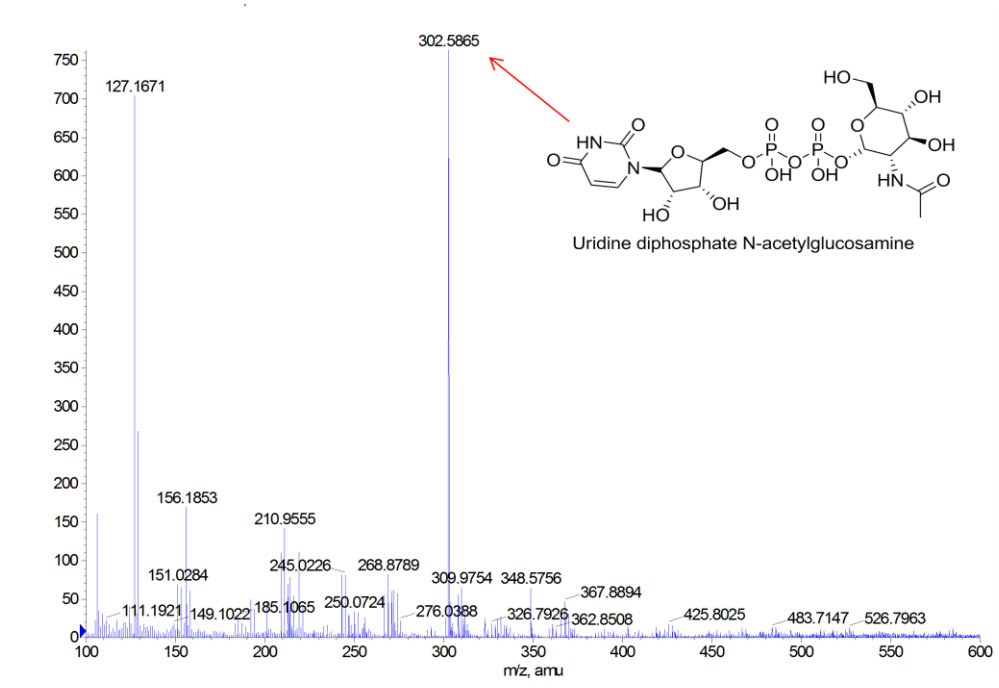


Figure 20. ESI spectrum of UDP-GlcNAc with TbMshA only.

This spectrum looks similar to the control in which it only has UDP-GlcNAc (dissolved in water), the noisy peaks at the background, as well as the sharp peak observed at 127 m/z for all the spectra are believed to be the fragments from the HEPES buffer (where the *in vitro* reactions were taken place). However, this spectrum cannot fully dispute the fact that UDP can be first hydrolyzed before transfer because the sample analyte was prepared by precipitating the protein (since some amount of ligand may be attached to the precipitating protein). A better approach to obtain the analyte could be to extract the solution form of the ligand from the ligand-bound TbMshA crystal directly to get a more definite and consistent results with the crystallographic data, unfortunately, this approach was unable to be followed in time.

3.12 The reaction sequence of UDP and I-1-P

Based on the ligand-bound TbMshA structure, UDP-GlcNAc is split into two separate pieces – UDP and GlcNAc (refinement supports this), in which they are securely stabilized by the neighboring residues in the structure (see the figure on p. 24). It is the liberation of UDP allows one to measure the enzyme activity of TbMshA using a coupled enzyme assay method, which is most commonly used for glycosyltransferase besides the traditional radio-labeling techniques (19,24). However, what is the reaction sequence of the donor-acceptor pair? The answer to this question is not clearly defined by the crystallographic structure alone. In the *in-vitro* reaction where UDP-GlcNAc is only with TbMshA, no detection of free UDP at all (see the figure on p. 41), and UDP is indeed present in the crystal structure but is not leaving yet (still in the active site

pocket). Previous experiments (data not shown) have had performed for TbMshA with only UDP-GlcNAc but not I-1-P showed a decreased slope but with less decline. It may indicate trace amount of UDP is releasing into the solution over time (the enzyme reaction was monitored over a period of 30 minutes). But, in the presence of I-1-P, the slope declines much faster (data not shown). The entry of acceptor seems to facilitate the leaving of UDP as one would predict from the crystal structure that there needs more room for the sugar transfer to take place.

3.13 E353Q impairs TbMshA activity and hinders product formation

The ligand-bound TbMshA structure shows Glu-353 is at a strategic spatial location, in which it is seemed to be mediating the interaction with UDP-GlcNAc. More importantly, as UDP-GlcNAc has split into two pieces (supported by the refinement, see the figure on p. 22), Glu-353 is still taking the spatial advantage over other nearby residues because it now becomes the "middle-man" of UDP and GlcNAc (see the figure on p. 24), allowing one to speculate its role as a true nucleophile in the catalysis. The mechanism proposed here (see the figure on p. 28) relies on the hypothesis that Glu-353 is serving as a nucleophile to initiate the attack to cleave the UDP-GlcNAc bond. Glu-353 uses its carboxyl (R-COO-) to attack on the C1' of the pyran ring of GlcNAc, allowing diphosphate to dissociate readily. The mutation of Glu-353 is simply a single-point mutation, in which the glutamate is mutated to glutamine through site-directed mutagenesis. Glutamate and glutamine are quite indistinguishable structurally because there's only one functional group difference (carboxylic group is changed to an amine

($R\text{-}CONH_2$)). If this carboxyl is the source of generating the nucleophile to initiate the attack, when the functional group is changed to amine, this source is completely blocked because amine ($\text{-}NH_2$) is less basic compared to glutamate, especially when the pH is at physiological range (pH 8.0). The weakening of the mutant TbMshA in its enzyme activity as well as the inability to produce its immediate product demonstrate that E353Q, after the glutamate is mutated to glutamine, is unable to serve as a nucleophile to facilitate the downstream reaction.

4. CONCLUSIONS

The structure of the precursor enzyme, MshA, in the mycothiol biosynthesis pathway in *Mtb*, is solved via x-ray crystallography to a resolution of 2.32 Å. The structure of the enzyme complexed with a ligand (UDP-GlcNAc) that is known to be the donor substrate is also solved using the same crystallographic technique, to a lesser resolution of 2.89 Å. Both structures form two monomers in the asymmetric unit with distinctive N- and C- terminal domains that are connected via a stretch of flexible loop, although such loop region is absent in the apo TbMshA structure. The tertiary structure of the monomer shows the MshA in *Mtb* is following a typical β/α/β Rossmann fold (GT-B fold) as seen in other retaining glycosyltransferase superfamily members. Compared to the ligand-bound structure, the apo structure of TbMshA is more open between the N- and C-terminal domain, leaving the connection loops between domains more mobile (thus higher disorderness in this region), however, this gap also opens the door to accept substrates. Once the substrate (or ligand) is bound, this gap somehow "closes up" in between the domains and be stabilized by the interactions with the ligand. This phenomenon is also observed in TbMshA's closest homologue in *Corynebacterium glutamicum* (CgMshA), however, the apo form of TbMshA has its C-terminal domain extend in lesser degree than that of CgMshA while the N-terminal domains in both structures stay rigid.

Ligand-bound TbMshA enables the visualization of several critical residues that are interacting with UDP-GlcNAc. Although based on the refinement of UDP-GlcNAc,

whose electron density supports the fragmentation into UDP and GlcNAc, the two fragments are still intact in its local interaction, except with GlcNAc moiety flipped 180^{o}, adopting its most favorable chair confirmation. Either the soaking or co-crystallization of the second substrate, I-1-P, to TbMshA tertiary complex is not successful, hence, no direct inference of mechanism can be drawn from the structure. However, the ligand-bound TbMshA crystal structure contains refined phosphate (which was the result from soaking as well), and the occupancy of this phosphate is in direct facing to the 1'-OH of GlcNAc, which indicates that this phosphate binding pocket maybe the cavity in which I-1-P will bind should it enter to the structure. Based on the location of residues that are in direct interactions with the UDP, GlcNAc and PO_4, a mechanism is drawn to illustrate the plausible sugar transfer. In addition, a critical residue, Glu-353, is indentified to be a nucleophile responsible for TbMshA catalysis.

The kinetics assays were carried out for both the wild-type and the mutant form (E353Q) of TbMshA. The K_m for wild-type TbMshA for UDP-GlcNAc and I-1-P are found to be 210 μM and 331 μM, respectively. The K_m for the mutant E353Q for UDP-GlcNAc and I-1-P are found to be 204 μM and 113 μM, respectively. The k_{cat} for the reaction catalyzed by wild-type TbMshA is calculated to be 0.17 ± 0.02 sec^{-1}, whereas for the mutant it is 0.04 ± 0.01 sec^{-1}, indicating the enzyme activity is greatly hampered by the mutant by at least 4-fold. k_{cat}/K_m for wild-type TbMshA is $3.5 \pm 1.1 * 10^3$ M^{-1} sec^{-1}, about ten times higher than that of the mutant, which has a k_{cat}/K_m of only $0.3 \pm 0.1 * 10^3$ M^{-1} sec^{-1}. In addition, mass spectrometry analysis through Electrospray Ionization (ESI) experiments show that the mutant is incapable of making the immediate product

(GlcNAc-Ins-P). Both the kinetic and mass spectrometry data support the hypothesis that mutation in Glu-353 causes the enzyme losing activity as well as preventing the formation of the product, thus, inhibiting the transfer reaction.

The findings here show that Glu-353 is needed for catalysis, and mutation to glutamine, which is structurally similar but functionally different, weakens the enzyme's activity and impede the transfer reaction (i.e. product formation). The major functional group of glutamate is the carboxylic group, which can serves as a nucleophile to activate the reaction whereas an amine group fails to do the same job in this case.

This thesis attempts to present a structural and functional analysis of the precursor enzyme, MshA, in the mycothiol biosynthesis pathway in *Mtb*. While the cure of TB remains as challenging as the eradication of *Mtb*, much more research work is needed to target this pathogen's own survival gear in the near future.

REFERENCES

1. World Health Organization (WHO). (2009). 10 Facts about Tuberculosis. WHO, http://www.who.int/features/factfiles/tuberculosis/en/index.html
2. TB Structural Genomics Consortium (TBSGC). (2011). http://www.webtb.org/
3. Newton, G. L., Arnold, K., Price, M. S., Sherrill, C., Delcardayre, S. B., Aharonowitz, Y., Cohen, G., Davies, J., Fahey, R. C., and Davis, C. (1996) *J Bacteriol* **178**, 1990-1995
4. Masip, L., Veeravalli, K., and Georgiou, G. (2006) *Antioxid Redox Signal* **8**, 753-762
5. Buchmeier, N. A., Newton, G. L., Koledin, T., and Fahey, R. C. (2003) *Molecular Microbiology* **47**, 1723-1732
6. Sareen, D., Newton, G. L., Fahey, R. C., and Buchmeier, N. A. (2003) *J Bacteriol* **185**, 6736-6740
7. Jothivasan, V. K., and Hamilton, C. J. (2008) *Nat Prod Rep* **25**, 1091-1117
8. Newton, G. L., Koledin, T., Gorovitz, B., Rawat, M., Fahey, R. C., and Av-Gay, Y. (2003) *J Bacteriol* **185**, 3476-3479
9. Newton, G. L., Av-Gay, Y., and Fahey, R. C. (2000) *J Bacteriol* **182**, 6958-6963
10. Sareen, D., Steffek, M., Newton, G. L., and Fahey, R. C. (2002) *Biochemistry* **41**, 6885-6890
11. Koledin, T., Newton, G. L., and Fahey, R. C. (2002) *Arch Microbiol* **178**, 331-337

12. Vilcheze, C., Av-Gay, Y., Attarian, R., Liu, Z., Hazbon, M. H., Colangeli, R., Chen, B., Liu, W., Alland, D., Sacchettini, J. C., and Jacobs, W. R., Jr. (2008) *Mol Microbiol* **69**, 1316-1329

13. Newton, G. L., Ta, P., Bzymek, K. P., and Fahey, R. C. (2006) *J Biol Chem* **281**, 33910-33920

14. Rose, N. L., Zheng, R. B., Pearcey, J., Zhou, R., Completo, G. C., and Lowary, T. L. (2008) *Carbohydr Res* **343**, 2130-2139

15. Otwinowski, Z., and Minor, W. (1997) *Processing of X-ray Diffraction Data Collected in Oscillation Mode*, Academic Press, New York

16. Rupp, B. (2009) *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*, Garland Science, New York

17. Collaborative Computational Project. (1994) *Acta Cryst* **D50**, 760-763

18. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. and Wilson, K. S. (2011) *Acta. Cryst.* **D67**, 235-242

19. Unligil, U. M., and Rini, J. M. (2000) *Curr Opin Struct Biol* **10**, 510-517

20. Newton, G. L., Ta, P., Bzymek, K. P., and Fahey, R. C. (2006) *J. Biol. Chem.* **281**, 33910-33920

21. Qasba, P. K., Ramakrishnan, B., and Boeggeman, E. (2005) *Trends Biochem Sci* **30**, 53-62

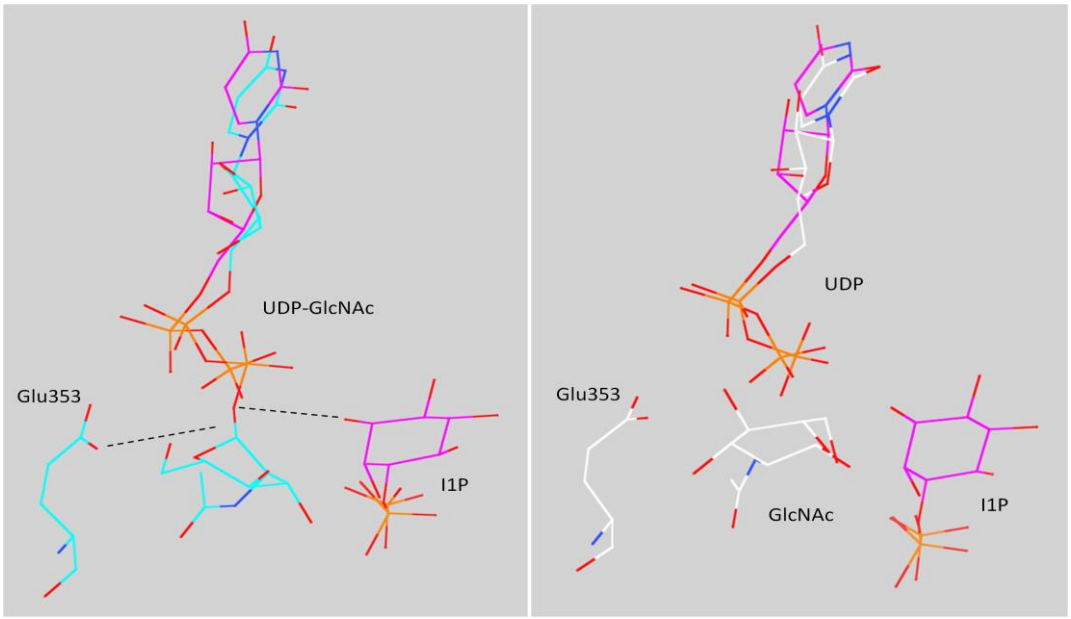
22. Vetting, M. W., Frantom, P. A., and Blanchard, J. S. (2008) *J Biol Chem* **283**, 15834-15844

23. Gosselin, S., Alhussaini, M., Streiff, M. B., Takabayashi, K., and Palcic, M. M. (1994) *Anal Biochem* **220**, 92-97

24. Lairson, L. L., Henrissat, B., Davies, G. J., and Withers, S. G. (2008) *Annu Rev Biochem* **77**, 521-555

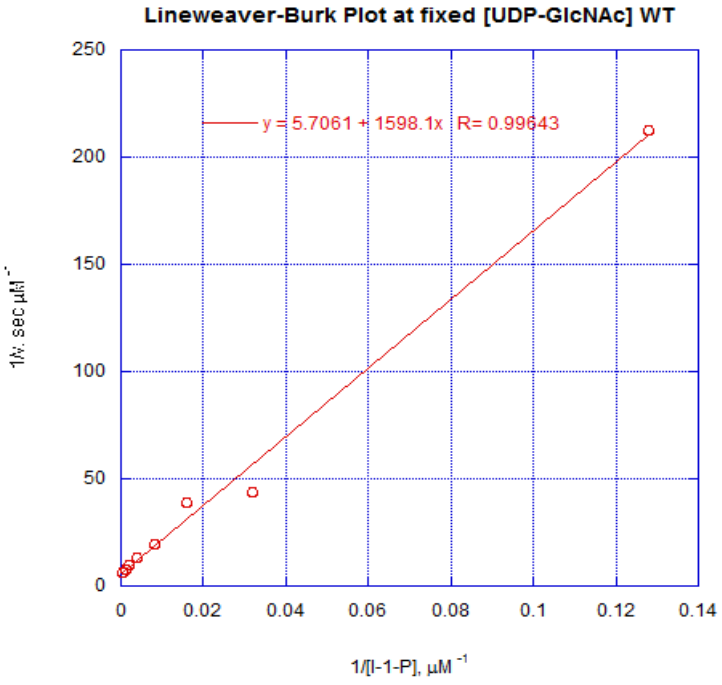
APPENDIX A

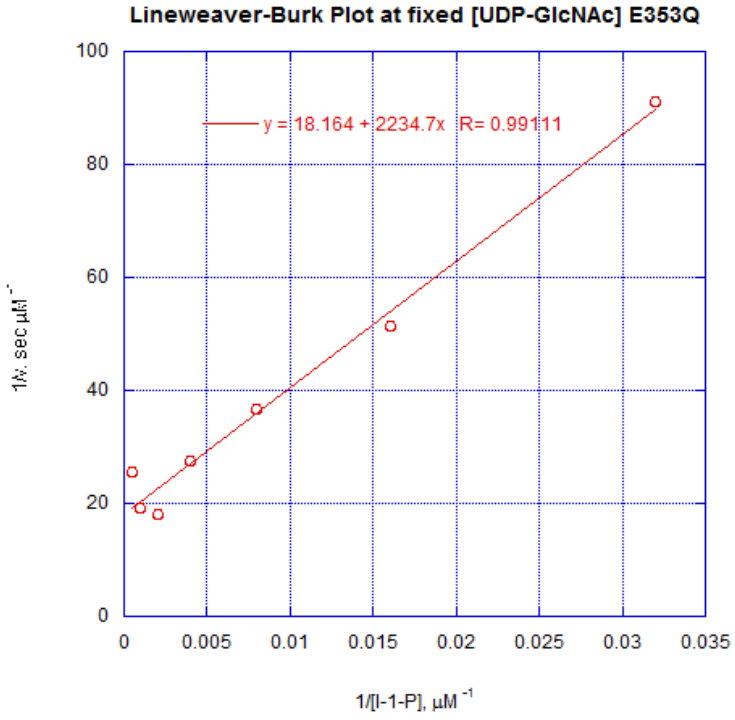
A) **B)**



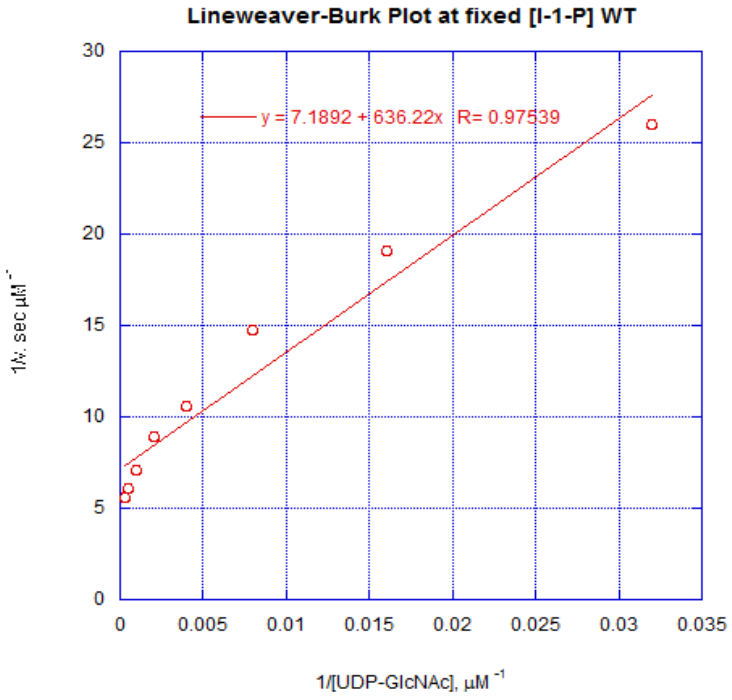
A composite figure generated by Chimera. Cyan: unhydrolyzed UDP-GlcNAc in TbMshA structure (previously refined); magenta: UDP in CgMshA structure (PDB code: 3c4v); white: hydrolyzed UDP-GlcNAc in TbMshA (current refinement). The UDP and PO_4 of TbMshA overlap with that of CgMshA.

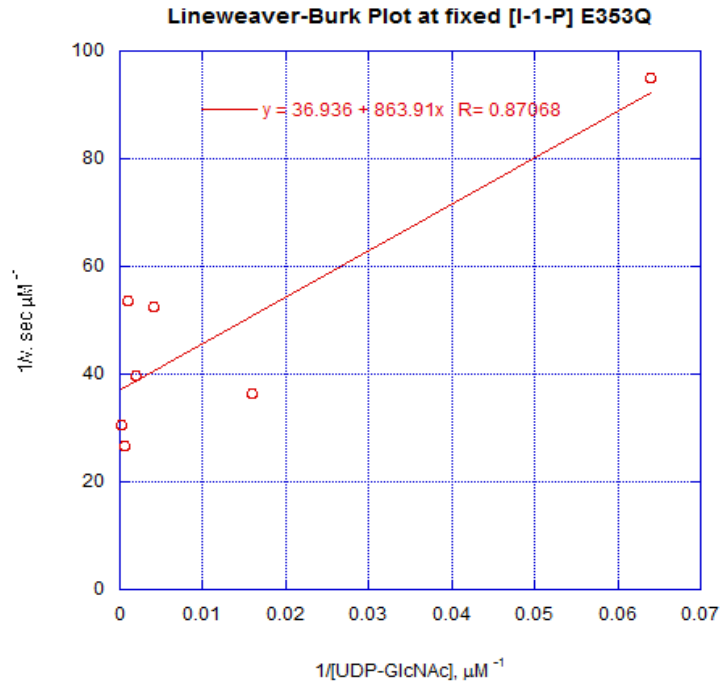
APPENDIX B-1


Lineweaver-Burk Plot at fixed [UDP-GlcNAc] WT
y = 5.7061 + 1598.1x R= 0.99643
1/v, sec µM -1
1/[I-1-P], µM -1
0
50
100
150
200
250
0.02
0.04
0.06
0.08
0.1
0.12
0.14

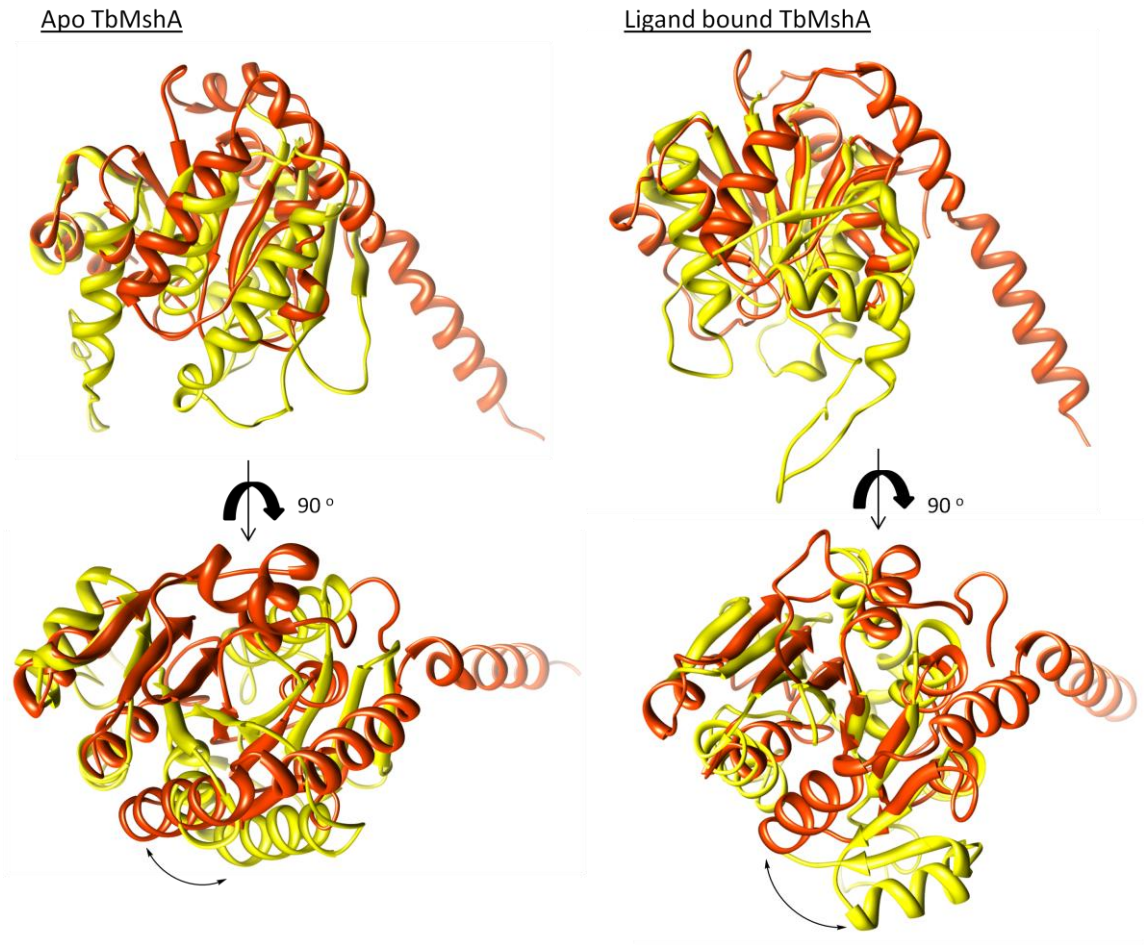

Lineweaver-Burk Plot at fixed [UDP-GlcNAc] E353Q
y = 18.164 + 2234.7x R= 0.99111
1/v, sec µM -1
1/[I-1-P], µM -1
0
20
40
60
80
100
0.005
0.01
0.015
0.02
0.025
0.03
0.035

APPENDIX B-2


Lineweaver-Burk Plot at fixed [I-1-P] WT
y = 7.1892 + 636.22x R= 0.97539
1/v: sec µM -1
1/[UDP-GlcNAc], µM -1
0
5
10
15
20
25
30
0
0.005
0.01
0.015
0.02
0.025
0.03
0.035


Lineweaver-Burk Plot at fixed [I-1-P] E353Q
y = 36.936 + 863.91x R= 0.87068
1/v: sec µM -1
1/[UDP-GlcNAc], µM -1
0
20
40
60
80
100
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07

APPENDIX C



Overlay of the N-domain (N-terminal to Pro-240, colored yellow) and C-domain (Gly-241 to C-terminal, colored orange red) for the apo and ligand-bound TbMshA structures with 90 degree of horizontal rotation shown at the bottom portion.

VITA

Name: Wan Wen Zhu

Address: Department of Chemistry
Texas A&M University
College Station, TX 77843-3012

Email Address: wzhu@mail.chem.tamu.com

Education: M.S., Chemistry, Texas A&M University, 2011

B.S., Chemistry, Polytechnic University, 2006